CRISPR-mediated epitope tagging of Sox proteins for ChIP-seq

by

Chamila Deshani Ranawakage
Student ID Number: 1208008

A dissertation submitted to the
Engineering course, Department of Engineering,
Graduate School of Engineering,
Kochi University of Technology,
Kochi, Japan

For the degree of
Doctor of Philosophy

Assessment Committee:
Supervisor: Prof. Yusuke Kamachi
Co-Supervisor: Prof. Takeshi Ohama
Co-Supervisor: Associate Prof. Sakae Horisawa
Prof. Seiji Tanaka
Prof. Masayuki Ike

September 2019

Abstract

ChIP-seq-based molecular analysis of a transcription factor (TF) has often been hampered by the lack of high-quality antibodies. Moreover, ChIP-seq analysis on closely related TF family members such as Sox proteins is hindered by the antibody cross-reactivity due to the high sequence similarity. Epitope tagging of TFs and subsequent use of epitope tag-specific antibodies provides a promising alternative for the requirement of "ChIP-grade" antibodies for each of the target protein of interest. Further, these epitope tags can be integrated into the gene of interest using the clustered regularly interspaced short palindromic repeats (CRISPR)/Cas9 genome editing tool. This study was conducted to develop methods related to CRISPR-mediated endogenous epitope tagging of *sox3* gene.

As the first objective, to select optimal epitope tag/antibody combinations suitable for successful ChIP experiments, a quite simple and relatively inexpensive approach was established to determine antibody affinity under IP/ChIP conditions termed HiBiT-qIP. By using this method, the performance of epitope tag/antibody combinations could be predicted quantitatively under ChIP assay conditions. The apparent affinities of interactions between five epitope tags, namely, FLAG, HA, PA, V5 and Ty1 and their cognate antibodies were determined. The use of epitope tags in multimeric form such as dimeric or trimeric form revealed a copy-number dependent increase in the apparent affinity and that improved the IP recovery significantly. As the second objective, an efficient epitope tag knock-in using CRISPR/Cas9 and long ssDNA (lssDNA) donor in zebrafish was achieved and thereby, transgenic zebrafish lines were established with epitope (FLAGx3, PAx3) tagged Sox3 proteins. A significantly high efficiency of precise and heritable integration of the composite of epitope tags (~200 bp in length) to the *sox3* locus was achieved. The use of lssDNA as a donor and its effective homology arm lengths were found to be a critical factor of the overall knock-in efficiency.

Table of Contents

Chapter 1.

General introduction

Sox (Sry-related HMG box) family transcription factors (TFs) play a major role as central coordinators of gene regulatory networks in the embryonic development processes such as cell differentiation and tissue development[1]. Our group has been studying these Sox TF-mediated transcriptional regulatory processes of early embryonic development using the zebrafish (*Danio rerio*) model, with a main focus on SoxB1 subfamily of TFs[2,3]. In the initial stages of zebrafish development, SoxB1 proteins have been found to regulate a number of genes by interacting with Pou5f1, a homolog of mammalian Oct4 in an analogous manner to Sox2-Oct4 partnership in ES cells[2,4]. Furthermore, in later stages of embryogenesis including the transition from epiblast to neuroectoderm, SoxB1 proteins appear to regulate different sets of target genes by pairing with other TFs[2,5,6]. Although the major players of these signaling pathways are known, there are many crucial details yet to be revealed in order to create a genome-wide picture of Sox TF-mediated gene regulatory network. For this purpose, chromatin immunoprecipitation followed by next-generation DNA sequencing (ChIP-seq) is a promising method, as it has already been used widely to map regulatory elements and analyze TF function throughout an entire genome[7,8,9].

However, the reliability of the results obtained through ChIP-seq highly depends on the affinity and the specificity of the antibodies used to capture the protein complexes[10]. Antibodies that show satisfying affinity and specificity during ChIP-seq are generally called "ChIP-grade" antibodies. Despite the popularity of ChIP-seq applications, the unavailability of ChIP-grade antibodies against each of the proteins of interest has hampered wider applications of ChIP-seq. Particularly, TF families like Sox proteins that share higher sequence similarities show antibody cross-reactivity and make it difficult to perform reliable ChIP-seq experiments[11].

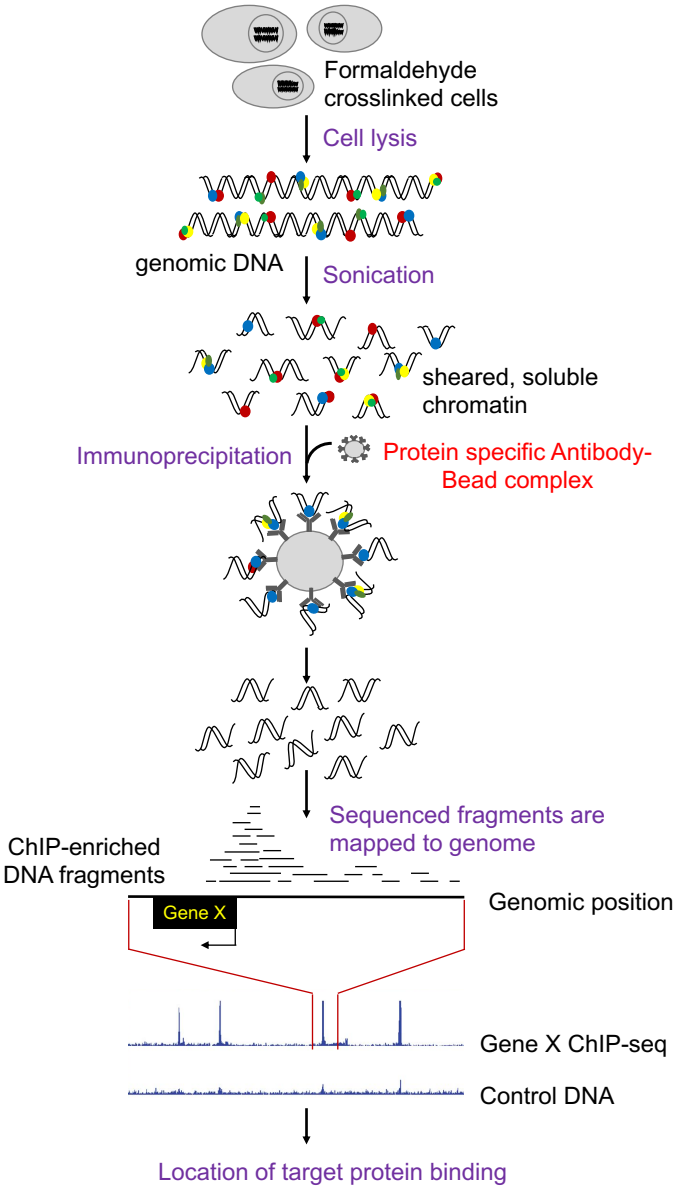


Figure 1-1. ChIP-seq work flow

Epitope tagging of the TFs and subsequent use of epitope tag-specific antibodies for ChIP-seq is an interesting strategy to avoid this problem. Further, this strategy attracts much attention due to the ability to adapt Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR)/Cas9 genome editing tool to epitope tag endogenous TFs and thus facilitates the expression of the tagged proteins at near-endogenous levels[12]. Amongst the

candidate epitope tags that can be used for ChIP-seq such as FLAG, HA and V5, the FLAG peptide is the most widely reported epitope tag in ChIP-seq experiments. An important consideration when selecting suitable epitope tags for ChIP is to avoid lysine or histidine residues within the amino acid sequence due to the fact that lysine residues are the primary targets for formaldehyde cross-linking. Thus, epitope tags that contain lysines will be functionally destroyed (at least partially) by formaldehyde cross-linking. For this reason, the FLAG tag has now raised controversies of its usage in ChIP-seq experiments. Another consideration when selecting epitope tags for ChIP-seq is their ability to obtain substantial enrichment over the control samples[13]. On the other hand, the performance of an epitope tag in a ChIP-seq experiment depends not only on the amino acid sequence of the epitope used but also on the quality of the anti-epitope antibody. Therefore, the selection of optimal epitope tag/antibody combinations is a prerequisite for successful ChIP experiments and remains to be evaluated.

1-1 Objectives of the study

- To select optimal epitope tag/antibody combinations suitable for successful ChIP experiments
- To generate transgenic zebrafish lines with epitope-tagged Sox proteins using CRISPR/Cas9 system

In this study, five different epitope tags such as FLAG, HA, PA, Ty1 and V5, and monoclonal antibodies against each of these epitope tags were selected based on their wide usage in research and the commercial availability. There was at least one candidate antibody clone to multiple candidate antibody clones against each epitope tag. Thus, the most suitable antibody clone had to be selected for further experiments, and that required the understanding of the antibody performance in advance especially, under the ChIP assay conditions.

Antibody performance can be evaluated by their specificity and/or affinity towards their cognate antigens. Here, I tried to develop a method to determine the antibody affinity or its quantitative determinant–dissociation constant (K_d), under IP conditions which is however challenging, mainly because the amount of precipitated protein during IP is often near or below the lowest limit of quantitative detection by Western blotting.

In chapter 2 of this dissertation, a simple and relatively inexpensive approach for determining the antibody K_d under IP conditions is introduced. This method was termed as HiBiT-qIP, which is short for "HiBiT-based quantitative immunoprecipitation". Through HiBiT-qIP, optimal epitope tag/antibody combinations that perform well under ChIP conditions could be selected. Next, I tried to tag zebrafish *sox3* gene endogenously with those epitope tags using CRISPR/Cas9-mediated knock-in (KI) approach, as described in chapter 3. Due to the reported low KI efficiency in zebrafish and the limited reports about successful long sequence insertions to the zebrafish genome, several factors such as the orientation of the single-stranded DNA (ssDNA), homology arm length, and symmetry or asymmetry of homology arms that would yield higher KI efficiency were examined.

1-2 References

1. Kamachi, Y. & Kondoh, H. Sox proteins: regulators of cell fate specification and differentiation. *Development* **140,** 4129–4144 (2013).
2. Okuda, Y., Ogura, E., Kondoh, H. & Kamachi, Y. B1 SOX Coordinate Cell Specification with Patterning and Morphogenesis in the Early Zebrafish Embryo. *PLoS Genet.* **6,** e1000936 (2010).
3. Okuda, Y. *et al.* Comparative genomic and expression analysis of group B1 sox genes in zebrafish indicates their diversification during vertebrate evolution. *Dev. Dyn.* **235,** 811–825 (2006).
4. Leichsenring, M., Maes, J., Mossner, R., Driever, W. & Onichtchouk, D. Pou5f1 Transcription Factor Controls Zygotic Gene Activation In Vertebrates. *Science (80-.).* **341,** 1005–1009 (2013).
5. Inoue, M. *et al.* PAX6 and SOX2-dependent regulation of the Sox2 enhancer N-3 involved in embryonic visual system development. *Genes to Cells* **12,** 1049–1061 (2007).
6. Kamachi, Y., Uchikawa, M. & Kondoh, H. Pairing SOX off with partners in the regulation of embryonic development. *Trends Genet.* **16,** 182–187 (2000).
7. Birney, E. *et al.* Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature* **447,** 799–816 (2007).
8. Dunham, I. *et al.* An integrated encyclopedia of DNA elements in the human genome. *Nature* **489,** 57–74 (2012).
9. Johnson, D. S., Mortazavi, A. & Myers, R. M. Protein-DNA Interactions. 1497–1503 (2007).
10. Landt, S. G. *et al.* ChIP-seq guidelines and practices of the ENCODE and modENCODE consortia. *Genome Res.* **22,** 1813–31 (2012).

11. Peter J. Park. ChIP-Seq: advantages and challenges of a maturing technology. *Nat. Rev. Genet.* **10,** 669–680 (2009).

12. Savic, D. *et al.* CETCh-seq: CRISPR epitope tagging ChIP-seq of DNA-binding proteins. *Genome Res.* **25,** 1581–1589 (2015).

13. Brinkman, A. B. & Stunnenberg, H. G. Strategies for epigenome analysis. in *Epigenomics* 3–18 (Springer Netherlands, 2009).

Chapter 2.

HiBiT-qIP, HiBiT-based quantitative immunoprecipitation, facilitates the determination of antibody affinity under immunoprecipitation conditions

2-1 Introduction

A broad range of research, diagnostic and therapeutic activities are inseparably linked to the use of antibodies for the enrichment, detection and quantitation of proteins and their modifications. The success of these procedures is highly dependent on the quality of the antibodies, which is critically determined by the affinity and specificity of the antibodies towards their cognate antigens. Although there are hundreds of thousands of commercially available antibodies, many of them have been poorly characterised and are thus inadequately reliable, which makes it difficult to find a suitable antibody for a specific application[1–6]. Immunoprecipitation (IP) is an immunological technique in which specific antibodies are used to enrich the target proteins or protein complexes from a protein mixture solution. IP has been extensively applied in many scientific fields to identify and study protein-protein and protein-DNA interactions[7]. Stringent assessment of the degree of sensitivity and specificity of an antibody in capturing its cognate antigen is required for a successful IP assay, particularly chromatin immunoprecipitation (ChIP)[2,8].

The sensitivity of an antibody-based assay is essentially determined by the binding affinity of the antibody to its cognate antigen. Thus, the measurement of the affinity of an antibody can predict its suitability for future IP experiments in advance. The dissociation constant (K_d) of an antigen-antibody interaction quantitatively defines its binding affinity. The most popular and widely used methods for determining K_d include enzyme-linked immunosorbent assay (ELISA)-based methods[9], surface plasmon resonance (SPR) biosensors[10,11] and the solution-based kinetic exclusion assay (KinExA)[12,13]. Each of these methods has its own inherent advantages and disadvantages[7,12,14], and different K_d values can

be obtained by the different types of immunological assays, as is described in the literature[12,15]. Because none of these methods reflects the IP protocol, it is difficult to predict the performance of an antibody during actual IP experiments, where antibody-antigen reactions are influenced by many factors. The critical parameters that affect the equilibrium constant are the ionic strength, pH, temperature, and the composition of ionic and nonionic detergents in the IP buffer[7,16,17]. Thus, determining apparent K_d values under particular IP conditions would be desirable. However, determining the K_d values of an antibody under IP conditions is challenging, mainly because the amount of precipitated protein is often near or below the lowest limit of quantitative detection by Western blotting, which is usually in the range of hundreds of picograms[18,19].

In this manuscript, we report a simple and relatively inexpensive approach for determining the antibody K_d under IP conditions by employing a quantitative NanoLuc-based HiBiT detection system. We call this method HiBiT-qIP, which is short for "HiBiT-based quantitative immunoprecipitation". The HiBiT system is based on the split luciferase complementation of two NanoLuc fragments. Specifically, a 1.3-kDa peptide (11 amino acids) is capable of producing bright luminescence through interaction with an 18-kDa polypeptide named Large BiT (LgBiT). During the development of the split luciferase complementation assay, a small peptide, which is called Small BiT (SmBiT) and has low affinity (K_d > 100 μM) to LgBiT, was initially adopted for the accurate measurement of protein interaction within cells[20]. In contrast, in the newly developed HiBiT system, the high-affinity (K_d = 0.7 nM) binding of a novel 11-amino acid High BiT (HiBiT) peptide to LgBiT efficiently forms a stable complex that acts as the active NanoLuc luciferase, which enables HiBiT to serve as a quantitative luminescent peptide tag[20,21]. Thus, tagging a protein of interest with the HiBiT peptide facilitates sensitive quantification of the amount of HiBiT-tagged protein, which makes it possible to measure protein amounts of less than 1

amol (e.g., 0.05 pg of a 50-kDa protein)[21–23]. Furthermore, a simple add-mix-read assay protocol of the HiBiT detection system enabled us to perform the IP-based equilibrium binding analysis more easily.

In the current study, we applied the HiBiT-qIP method to evaluate monoclonal antibodies against epitope tags that are widely utilised in immunoprecipitation. Epitope tagging of a target protein with a short peptide and subsequent use of epitope-specific antibodies to immunoprecipitate the tagged protein is a promising strategy for circumventing the lack of antibodies against target proteins[24–27]. This strategy is gaining popularity because an epitope tag can now be introduced into an endogenous target protein by adapting the clustered regularly interspaced short palindromic repeats (CRISPR)/Cas9 genome editing tool to express the tagged proteins at near-endogenous levels[27–30]. Here, we examined the affinities of monoclonal antibodies against the epitope tags of FLAG[31,32], HA[33] and V5[34], PA[35] and Ty1[36] because little information on their K_d values is currently available in spite of their widespread usage[26,37]. The PA tag was examined because it was recently reported that the rat monoclonal antibody NZ-1 against human podoplanin can be used as a high-affinity tagging system[35]. The Ty1 tag reportedly exhibits high performance in ChIP and was thus also included in our analysis[38].

The performance of an epitope tag in an IP experiment depends not only on the amino acid sequence of the epitope tag used but also substantially on the quality of the anti-epitope tag antibody. Furthermore, a number of monoclonal clones for some epitope tags, such as FLAG, have been developed and are commercially available. Therefore, the selection of the optimal epitope tag/antibody combination is a prerequisite for truly successful IP experiments, but such selection has rarely been performed. Moreover, epitope tags have often been used as multimerised forms to increase the efficiency of IP experiments, but their effects have not been quantitatively studied. Here, we aimed to evaluate the optimal epitope tag/antibody

combinations suitable for IP and the effects of tag multimerisation by developing the HiBiT-based quantitative immunoprecipitation assay (HiBiT-qIP) and using this assay to determine the apparent K_d values of various combinations. As we compared a collection of epitope tags in combination with commercially available monoclonal antibodies, the findings of this study will constitute a valuable resource for future IP-related experiments.

2-2 Results

2-2-1 Design of an assay for the determination of antibody affinity using the HiBiT system.

The NanoLuc-based HiBiT system is an accurate and sensitive protein quantification approach due to the linearity and stability of the luminescence signal generated by HiBiT/LgBiT, namely, the reconstituted NanoLuc luciferase[21]. By tagging a protein of interest with the HiBiT peptide, its amount can be easily and accurately quantified using the HiBiT detection reagent containing LgBiT and the luciferase substrate furimazine[21–23]. Employing this system, we designed an assay to evaluate the suitability of an antibody for IP by determining the antibody dissociation constant K_d under specific IP reaction conditions. In the current study, we applied the HiBiT-qIP assay to determine the K_d values of monoclonal antibodies against the epitope tags FLAG, HA, V5, PA and Ty1, which were selected based on their wide usage in research and the commercial availability of corresponding monoclonal antibodies (Tables 2-1 and 2-2).

Epitope tag	Sequence
FLAG	**DYKDDDDK**
FLAGx3	**DYKDDDDK**G**DYKDDDDK**I**DYKDDDDK**
HA	**YPYDVPDYA**
HAx3	**YPYDVPDYA**G**YPYDVPDYA**G**YPYDVPDYA**
V5	**GKPIPNPLLGLDST**
V5x2	**GKPIPNPLLGLDST**G**GKPIPNPLLGLDST**
V5x3	**GKPIPNPLLGLDST**G**GKPIPNPLLGLDST**G**GKPIPNPLLGLDST**
PA	**GVAMPGAEDDVV**
PAx2	**GVAMPGAEDDVV**G**GVAMPGAEDDVV**
PAx3	**GVAMPGAEDDVV**G**GVAMPGAEDDVV**TR**GVAMPGAEDDVV**
Ty1	**EVHTNQDPLD**
Ty1x2	**EVHTNQDPLD**A**EVHTNQDPLD**
Ty1x3	**EVHTNQDPLD**A**EVHTNQDPLD**TR**EVHTNQDPLD**

Table 2-1. Evaluated epitope tag sequences.

Antibody type	Supplier	Clone	Species	IgG subclass	Bead type [a]	Tag	K_d (nM)	95% Confidence interval of K_d (nM)
Anti-FLAG	Sigma	M2	Mouse	IgG1	α-mouse IgG beads	FLAG	0.76 [b]	0.44-1.3
						FLAGx3	0.21	0.12-0.37
	Wako	IE6	Mouse	IgG2b	α-mouse IgG beads	FLAG	1.8	1.3-2.6
						FLAGx3	0.50	0.40-0.62
	MBL	FLA-1	Mouse	IgG2a κ	α-mouse IgG beads	FLAG	1.3	0.82-2.0
						FLAGx3	0.33	0.22-0.50
	BioLegend	L5	Rat	IgG2a	α-rat IgG beads	FLAG	0.44	0.27-0.72
						FLAGx3	0.16	0.097-0.20
Anti-HA	Roche	3F10	Rat	IgG1	α-rat IgG beads	HA	0.38	0.22-0.70
						HAx3	0.067	0.035-0.12
	Wako	4B2	Mouse	IgG2b	α-mouse IgG beads	HA	6.6 [b]	5.5-8.0
						HAx3	0.88	0.50-1.5
Anti-V5	Sigma	V5-10	Mouse	IgG1	α-mouse IgG beads	V5	0.59	0.37-0.95
						V5x2	0.36	0.22-0.58
						V5x3	0.23	0.15-0.35
	Wako	6F5	Mouse	IgG2b	α-mouse IgG beads	V5	0.42	0.22-0.77
						V5x2	0.28	0.16-0.47
						V5x3	0.28	0.18-0.44
Anti-PA	Wako	NZ-1	Rat	IgG2a	α-rat IgG beads	PA	0.65 [b]	0.38-1.1
						PAx2	0.38	0.26-0.54
						PAx3	0.27	0.17-0.42
Anti-Ty1	Sigma	BB2	Mouse	IgG1	α-mouse IgG beads	Ty1	0.39 [b]	0.22-0.65
						Ty1x2	0.29	0.17-0.48
						Ty1x3	0.20	0.12-0.34

Table 2-2. Detailed overview of characterised antibodies.

[a] Anti-IgG magnetic beads used to capture the antibodies.

[b] K_d values obtained by combining data from two independent experiments.

These epitope tags, in their monomeric, dimeric (x2) or trimeric (x3) form, were first fused to the GST protein and the HiBiT peptide (Fig. 2-1Aa, Fig. 2-2, Table 2-1). The epitope-tagged GST proteins were then expressed in and purified from *E. coli* to near homogeneity. The purified proteins were separated using SDS-polyacrylamide gels, stained with Coomassie Brilliant Blue G-250 and quantified based on the infrared fluorescence of Coomassie blue[39] (Fig. 2-3A, B). We confirmed the full-length protein bands using the Nano-Glo HiBiT Blotting System[22,23] (Fig. 2-3C) and quantified only the intact proteins.

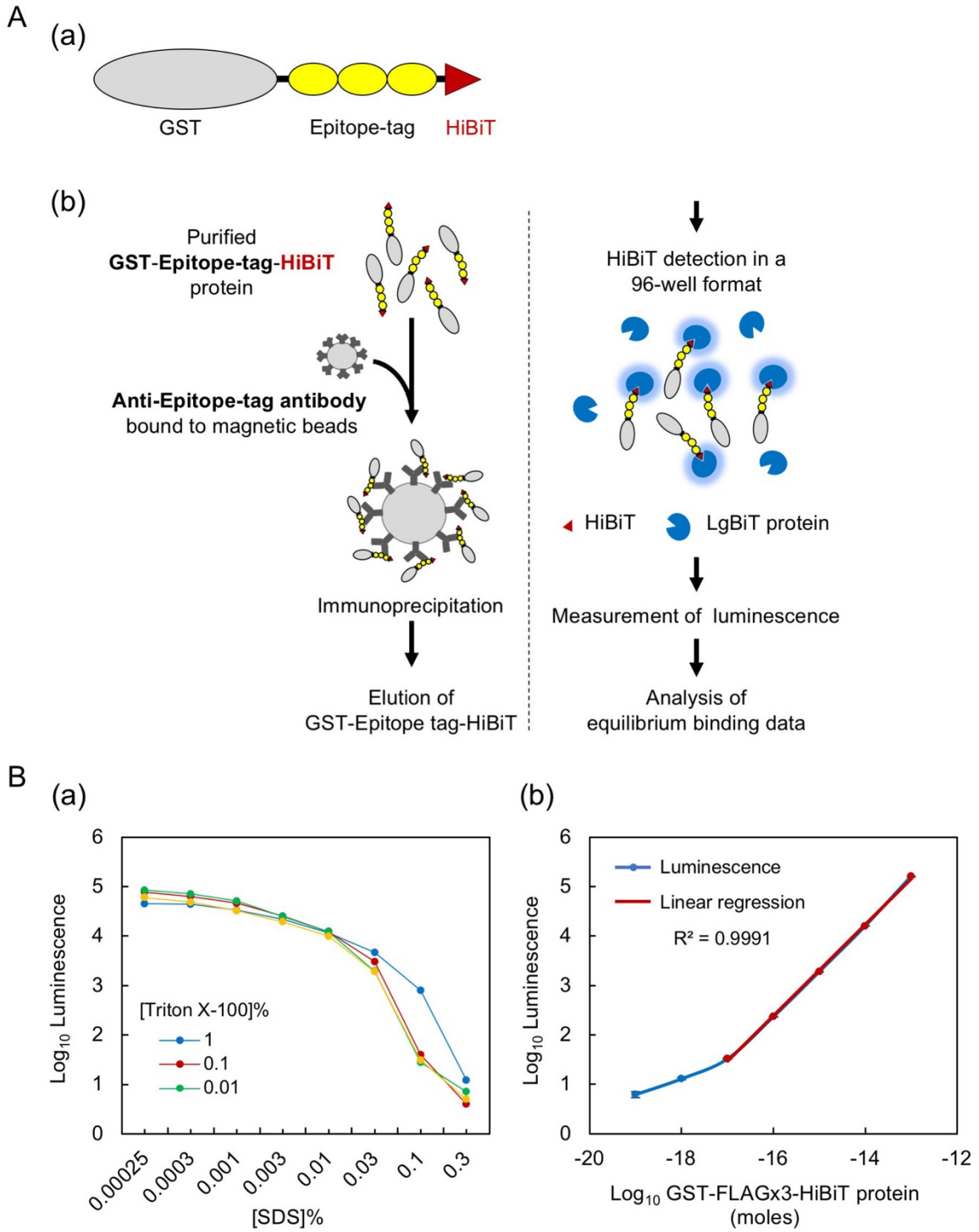


Figure 2-1. HiBiT-based quantitative immunoprecipitation. (**A**) Design of the assay. (**a**) Schematic representation of the GST-epitope tag-HiBiT fusion protein. The coding region of the GST gene is C-terminally fused to the FLAG, HA, V5, PA or Ty1 epitope tags in their monomeric, dimeric or trimeric form and the HiBiT peptide, which is placed in the most C-terminal position. In this panel, the trimeric form of the epitope tags is shown as an example; the tags are not drawn to scale. (**b**) Illustration showing the main steps of the HiBiT-qIP assay and the principle of HiBiT detection. The details are provided in the main text. (**B**) HiBiT protein quantitation in the presence of SDS. (**a**) Effect of SDS and Triton X-100 on the HiBiT solution assay. To examine the effects of SDS on the enzymatic activity

of reconstituted NanoLuc, 0.2 ng of GST-FLAGx3-HiBiT protein was included in 20 µL of PBS containing one of a series of concentrations of SDS (0.00025 to 0.3%), and the luminescence was measured after the addition of HiBiT detection reagents. The optimal Triton X-100 concentration for quenching the SDS effect was determined by adding Triton X-100 at three different concentrations, as indicated. (**b**) Linearity of the luminescence generated by HiBiT-LgBiT under our assay conditions. A tenfold dilution series of GST-FLAGx3-HiBiT protein (3.3 fg [10^{-19} moles] to 3.3 ng [10^{-13} moles]) in 20 µL of PBS containing 0.001% SDS, 0.01% BSA and 0.1% Triton X-100 was used in the HiBiT solution assay.

Varying amounts of the purified epitope-tagged GST protein were then mixed with a fixed amount of cognate monoclonal antibody immobilised on anti-IgG magnetic beads in a stringent IP buffer, which has been extensively used as the buffer for radio-immunoprecipitation assays (RIPAs)[7,40,41]. Importantly, the amount of antibody used during IP was optimised to maintain the concentration close to, or lower than, the K_d of each antibody, as suggested for standard binding assays[42]. The IP mixtures were incubated overnight at 4 ºC, during which time the binding reaction between the antigen and antibody was assumed to reach equilibrium because most IP reactions reportedly reach the plateau phase within a few hours[16,43,44]. After overnight incubation of the IP mixtures, the unbound proteins were washed away, and the amount of bound epitope-tagged GST protein was determined by measuring the luminescence signal derived by the HiBiT/LgBiT complex (Fig. 2-1Ab). A saturation curve of bound GST as a function of free GST was plotted by fitting the data to the binding model mentioned in the methodology section, and the K_d values were determined. For all K_d determinations, error graphs were plotted, and the 95% confidence intervals were calculated. We consider the obtained K_d values as "apparent" K_d values under our IP conditions. The "apparent" K_d values take into consideration factors such as antibody valency, steric hindrance and the mode of antibody immobilisation[45,46]. The apparent K_d values thus may not be identical to true K_d values that would be obtained through an ideal assay using a completely homogeneous solution.

```
(A)
1<----------FLAG--------------->---------<----------TEV------------>---------<------------------Bio
tag------------------------------>---------GAT TAT AAA GAC GAC GAT GAT AAG ACG CGT GAG AAC CTG TAC TTC CAG GGT GGA GGG GGA CTC AAT GAC ATT TTT GAA
GCT CAA AAG ATC GAG TGG CAC GAG GGC GCT
 D  Y  K  D  D  D  D  K  T  R  E  N  L  Y  F  Q  G  G  G  G  L  N  D  I  F  E  A  Q  K  I  E  W  H  E  G  A
<-----------------HiBiT------------------->-144
GTG AGC GGC TGG CGG CTG TTC AAG AAG ATT AGC TAA
 V  S  G  W  R  L  F  K  K  I  S  *
(B)
1<----------------------------------------------FLAGx3----------------------------------------------->---------<----------TEV------------>----
GAT TAT AAA GAC GAC GAT GAC AAA GGA GAC TAC AAG GAC GAT GAC GAC AAA ATT GAT TAC AAG GAC GAT GAT GAT AAG ACG CGT GAG AAC CTG TAC TTC CAG GGT GGA D
Y  K  D  D  D  D  K  G  D  Y  K  D  D  D  D  K  I  D  Y  K  D  D  D  D  K  T  R  E  N  L  Y  F  Q  G  G
-----<------------------Bio tag------------------------------>---------<-----------------HiBiT------------------->-198
GGG GGA CTC AAT GAC ATT TTT GAA GCT CAA AAG ATC GAG TGG CAC GAG GGC GCT GTG AGC GGC TGG CGG CTG TTC AAG AAG ATT AGC TAA
 G  G  L  N  D  I  F  E  A  Q  K  I  E  W  H  E  G  A  V  S  G  W  R  L  F  K  K  I  S  *
(C)
1<--------------HA---------------->---------<----------TEV------------>---------<------------------Bio
tag------------------------------>----TAT CCG TAT GAT GTG CCT GAT TAC GCT ACG CGT GAG AAC CTG TAC TTC CAG GGT GGA GGG GGA CTC AAT GAC ATT TTT GAA
GCT CAA AAG ATC GAG TGG CAC GAG GGC Y  P  Y  D  V  P  D  Y  A  T  R  E  N  L  Y  F  Q  G  G  G  G  L  N  D  I  F  E  A  Q  K  I  E
W  H  E  G  -----<----------------HiBiT------------------->-147
GCT GTG AGC GGC TGG CGG CTG TTC AAG AAG ATT AGC TAA
 A  V  S  G  W  R  L  F  K  K  I  S  *
(D)
1<-----------------------------------------------------HAx3------------------------------------------------------>---------<----------TEV-----
TAT CCG TAT GAT GTG CCT GAT TAC GCT GGC TAC CCC TAC GAT GTC CCA GAC TAC GCC GGA TAC CCT TAT GAC GTT CCA GAC TAT GCA ACG CGT GAG AAC CTG TAC TTC
 Y  P  Y  D  V  P  D  Y  A  G  Y  P  Y  D  V  P  D  Y  A  G  Y  P  Y  D  V  P  D  Y  A  T  R  E  N  L  Y  F
-------><---------<------------------Bio tag------------------------------>---------<----------------HiBiT------------------->-147
CAG GGT GGA GGG GGA CTC AAT GAC ATT TTT GAA GCT CAA AAG ATC GAG TGG CAC GAG GGC GCT GTG AGC GGC TGG CGG CTG TTC AAG AAG ATT AGC TAA
 Q  G  G  G  G  L  N  D  I  F  E  A  Q  K  I  E  W  H  E  G  A  V  S  G  W  R  L  F  K  K  I  S  *
(E)
1<--------------------PA----------------------->---------<----------TEV------------>---------<------------------Bio
tag------------------------GGC GTT GCC ATG CCA GGT GCC GAA GAT GAT GTG GTG ACG CGT GAG AAC CTG TAC TTC CAG GGT GGA GGG GGA CTC AAT GAC ATT TTT GAA
GCT CAA AAG ATC GAG TGG G  V  A  M  P  G  A  E  D  D  V  V  T  R  E  N  L  Y  F  Q  G  G  G  G  L  N  D  I  F  E  A  Q  K  I  E
W  ------->--------<-----------------HiBiT------------------->-156
CAC GAG GGC GCT GTG AGC GGC TGG CGG CTG TTC AAG AAG ATT AGC TAA
 H  E  G  A  V  S  G  W  R  L  F  K  K  I  S  *
```

Figure 2-2: Nucleotide and amino acid sequences of composite epitope tags

```
(F)
1<-----------------------------------------------PAx2-----------------------------------------------><---------<----------TEV------------>--------
GGG GTT GCA ATG CCT GGT GCT GAG GAT GAT GTC GTG GGC GGA GTG GCT ATG CCA GGC GCG GAA GAC GAC GTA GTG ACG CGT GAG AAC CTG TAC TTC CAG GGT GGA GGG G
V  A  M  P  G  A  E  D  D  V  V  G  G  V  A  M  P  G  A  E  D  D  V  V  T  R  E  N  L  Y  F  Q  G  G  G
-<-----------------Bio tag--------------------------------><---------<-----------------HiBiT------------------->-195
GGA CTC AAT GAC ATT TTT GAA GCT CAA AAG ATC GAG TGG CAC GAG GGC GCT GTG AGC GGC TGG CGG CTG TTC AAG AAG ATT AGC TAA
 G  L  N  D  I  F  E  A  Q  K  I  E  W  H  E  G  A  V  S  G  W  R  L  F  K  K  I  S  *
(G)
1<-------------------------------------------------------------------PAx3-------------------------------------------------------------------
GGG GTT GCA ATG CCT GGT GCT GAG GAT GAT GTC GTG GGC GGA GTG GCT ATG CCA GGC GCG GAA GAC GAC GTA GTG ACG CGC GGC GTT GCC ATG CCA GGT GCC GAA GAT
 G  V  A  M  P  G  A  E  D  D  V  V  G  G  V  A  M  P  G  A  E  D  D  V  V  T  R  G  V  A  M  P  G  A  E  D
---------->---------<----------TEV------------><---------<-----------------------Bio tag--------------------------><---------<----------------
GAT GTG GTG ACG CGT GAG AAC CTG TAC TTC CAG GGT GGA GGG GGA CTC AAT GAC ATT TTT GAA GCT CAA AAG ATC GAG TGG CAC GAG GGC GCT GTG AGC GGC TGG CGG
 D  V  V  T  R  E  N  L  Y  F  Q  G  G  G  G  L  N  D  I  F  E  A  Q  K  I  E  W  H  E  G  A  V  S  G  W  R
HiBiT--------------------237
CTG TTC AAG AAG ATT AGC TAA
 L  F  K  K  I  S  *
(H)
1<--------------------V5------------------------------><---------<----------TEV------------><---------<-------------------Bio
tag----------------GGT AAG CCC ATT CCA AAC CCA CTC CTG GGG TTA GAC TCT ACC ACG CGT GAG AAC CTG TAC TTC CAG GGT GGA GGG GGA CTC AAT GAC ATT TTT GAA
GCT CAA AAG ATC
 G  K  P  I  P  N  P  L  L  G  L  D  S  T  T  R  E  N  L  Y  F  Q  G  G  G  G  L  N  D  I  F  E  A  Q  K  I
--------------> --------<-----------------HiBiT------------------->-162
GAG TGG CAC GAG GGC GCT GTG AGC GGC TGG CGG CTG TTC AAG AAG ATT AGC TAA
 E  W  H  E  G  A  V  S  G  W  R  L  F  K  K  I  S  *
(I)
1<-----------------------------------------------V5x2-----------------------------------------------------------><---------<----------TEV-----
GGT AAG CCC ATT CCA AAC CCA CTC CTG GGG TTA GAC TCT ACC GGA GGC AAA CCG ATC CCT AAT CCT CTT TTG GGA CTG GAT AGC ACA ACG CGT GAG AAC CTG TAC TTC G
K  P  I  P  N  P  L  L  G  L  D  S  T  G  G  K  P  I  P  N  P  L  L  G  L  D  S  T  T  R  E  N  L  Y  F
------->---------<-----------------Bio tag--------------------------------><---------<-----------------HiBiT------------------->-207
CAG GGT GGA GGG GGA CTC AAT GAC ATT TTT GAA GCT CAA AAG ATC GAG TGG CAC GAG GGC GCT GTG AGC GGC TGG CGG CTG TTC AAG AAG ATT AGC TAA
 Q  G  G  G  G  L  N  D  I  F  E  A  Q  K  I  E  W  H  E  G  A  V  S  G  W  R  L  F  K  K  I  S  *
```

Figure 2-2: Nucleotide and amino acid sequences of composite epitope tags

```
(J)
1<-------------------------------------------------------------------------V5x3-------------------------------------------------------------------------
GGA AAG CCA ATC CCG AAT CCC CTT CTT GGA CTC GAC TCC ACA GGA GGT AAG CCC ATT CCA AAC CCA CTC CTG GGG TTA GAC TCT ACC GGA GGC AAA CCG ATC CCT AAT G
K  P  I  P  N  P  L  L  G  L  D  S  T  G  G  K  P  I  P  N  P  L  L  G  L  D  S  T  G  G  K  P  I  P  N
------------------------------>---------<----------TEV------------>---------<-----------------Bio
tag---------------------------------->--------CCT CTT TTG GGA CTG GAT AGC ACA ACG CGT GAG AAC CTG TAC TTC CAG GGT GGA GGG GGA CTC AAT GAC ATT TTT
GAA GCT CAA AAG ATC GAG TGG CAC GAG GGC GCT
 P  L  L  G  L  D  S  T  T  R  E  N  L  Y  F  Q  G  G  G  G  L  N  D  I  F  E  A  Q  K  I  E  W  H  E  G  A
-<----------------HiBiT------------------->-252
GTG AGC GGC TGG CGG CTG TTC AAG AAG ATT AGC TAA
 V  S  G  W  R  L  F  K  K  I  S  *
(K)
1<--------------Ty1------------------->---------<----------TEV------------>---------<------------------Bio
tag-------------------------------GAG GTG CAC ACT AAT CAA GAT CCT CTG GAC ACG CGT GAG AAC CTG TAC TTC CAG GGT GGA GGG GGA CTC AAT GAC ATT TTT GAA
GCT CAA AAG ATC GAG TGG CAC GAG E  V  H  T  N  Q  D  P  L  D  T  R  E  N  L  Y  F  Q  G  G  G  G  L  N  D  I  F  E  A  Q  K  I  E
W  H  E
>--------<-----------------HiBiT------------------->-150
GGC GCT GTG AGC GGC TGG CGG CTG TTC AAG AAG ATT AGC TAA
 G  A  V  S  G  W  R  L  F  K  K  I  S  *
(L)
1<-------------------------------------Ty1X2------------------------------------->---------<----------TEV------------>---------<--------------
GAG GTG CAT ACA AAT CAG GAC CCT CTC GAT GCT GAA GTC CAC ACC AAC CAA GAT CCA CTG GAC ACG CGT GAG AAC CTG TAC TTC CAG GGT GGA GGG GGA CTC AAT GAC E
V  H  T  N  Q  D  P  L  D  A  E  V  H  T  N  Q  D  P  L  D  T  R  E  N  L  Y  F  Q  G  G  G  G  L  N  D
-----Biotag------------------------------->---------<-----------------HiBiT------------------->-183
ATT TTT GAA GCT CAA AAG ATC GAG TGG CAC GAG GGC GCT GTG AGC GGC TGG CGG CTG TTC AAG AAG ATT AGC TAA
 I  F  E  A  Q  K  I  E  W  H  E  G  A  V  S  G  W  R  L  F  K  K  I  S  *
(M)
1<-------------------------------------------------------------Ty1X3------------------------------------------------------------------>---------<--
GAG GTG CAT ACA AAT CAG GAC CCT CTC GAT GCT GAA GTC CAC ACC AAC CAA GAT CCA CTG GAC ACG CGC GAG GTG CAC ACT AAT CAA GAT CCT CTG GAC ACG CGT GAG
 E  V  H  T  N  Q  D  P  L  D  A  E  V  H  T  N  Q  D  P  L  D  T  R  E  V  H  T  N  Q  D  P  L  D  T  R  E
--------TEV------------>---------<-------------------Biotag------------------------------->---------<-----------------HiBiT-------------------
AAC CTG TAC TTC CAG GGT GGA GGG GGA CTC AAT GAC ATT TTT GAA GCT CAA AAG ATC GAG TGG CAC GAG GGC GCT GTG AGC GGC TGG CGG CTG TTC AAG AAG ATT AGC
 N  L  Y  F  Q  G  G  G  G  L  N  D  I  F  E  A  Q  K  I  E  W  H  E  G  A  V  S  G  W  R  L  F  K  K  I  S  >-219
TAA
 *
```

Figure 2-2: Nucleotide and amino acid sequences of composite epitope tags. The amino acid sequences of the composite epitope tags are shown with the corresponding nucleotide sequences. Each nucleotide sequence is proceeded by the *Xho*I site sequence CTCGAG and followed by the *Xba*I-*Not*I linker sequence TCTAGAGCGGCCGC for cloning into pGEX-6P-1. TEV, the TEV protease cleavage site; Bio tag, the biotin ligase recognition site.

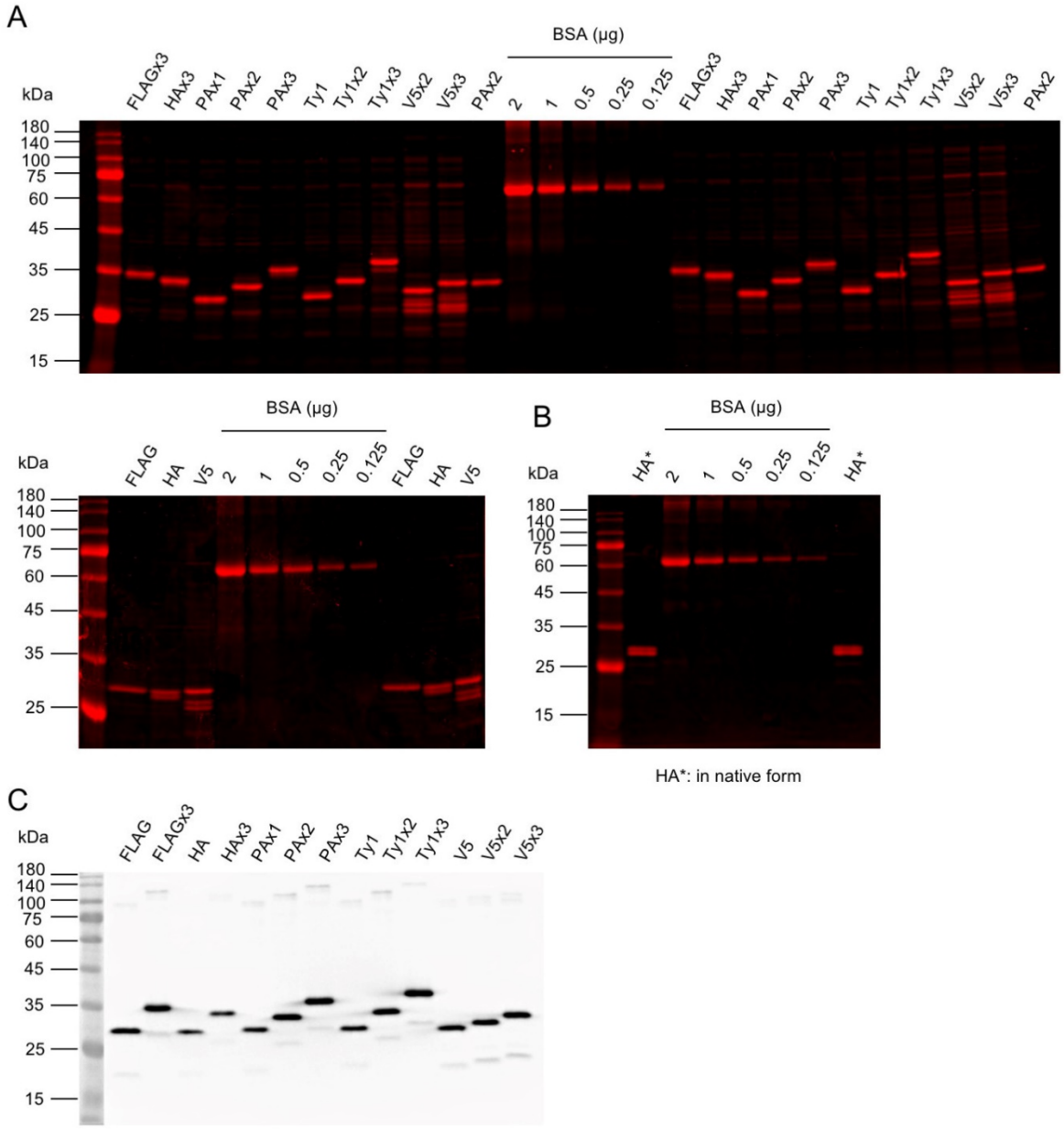


Figure 2-3. GST-epitope tag fusion protein quantification. (A) SDS polyacrylamide gel images of the GST proteins fused with monomeric, dimeric and trimeric forms of the epitope tags, which were prepared in denatured form. Each GST fusion was loaded in duplicate along with BSA standards for quantification. The gels were stained with Coomassie Brilliant Blue G-250 and detected using near-infrared fluorescence. (B) SDS polyacrylamide gel image of the GST protein fused with a monomeric form of the HA epitope tag, which was prepared in native form. (C) HiBiT blot detection of the series of tagged GST proteins to confirm retention of the C-terminal HiBiT tag.

2-2-2 HiBiT protein quantitation can be performed in the presence of residual SDS.

When using the HiBiT system for IP experiments, one should consider the effect of the residual SDS derived from the IP elution buffer on the interaction between HiBiT and LgBiT. Therefore, we first examined the effects of SDS on the HiBiT solution assay by measuring the luminescence values in the presence of varying concentrations of SDS in the sample solution. We also sought to determine the optimal concentration of Triton X-100 that could effectively quench the disruptive effect of SDS. When we used 0.2 ng of the purified GST-FLAGx3-HiBiT protein in the HiBiT solution assay, SDS clearly inhibited the interaction between HiBiT and LgBiT, even at low concentrations (Fig. 2-1Ba). The results also showed that 1% Triton X-100 exerted an SDS-quenching effect in the presence of >0.01% SDS, as expected, but slightly inhibited the HiBiT solution assay in the presence of <0.01% SDS. At these lower SDS concentrations, moderate concentrations of Triton X-100 exhibited the SDS-quenching effect. Based on these observations, we adjusted the final SDS concentration to 0.001% and added 0.1% Triton X-100 to the assay samples in the subsequent experiments. In addition, to minimise the SDS concentration in the assay samples, we used IP elution buffer containing 0.1% SDS and 25 mM DTT.

We then confirmed the linearity of the luminescence generated by HiBiT/LgBiT under the above conditions. Specifically, a tenfold dilution series was prepared starting from 3.3 ng of GST-FLAGx3-HiBiT with phosphate buffered saline (PBS) containing 0.01% bovine serum albumin (BSA) in addition to 0.1% TritonX-100 and 0.001% SDS. In the presence of saturating LgBiT in the HiBiT assay reagent solution, GST-FLAGx3-HiBiT produced luminescent signals that were linearly correlated to the protein amounts (shown in red line in Fig. 1Bb), with a lower limit of approximately 0.33 pg (0.01 fmol).

2-2-3 The K_d values varied considerably among monoclonal antibody clones.

We first determined the K_d values of various monoclonal antibodies against the epitope tags FLAG, HA, V5, PA and Ty1, which are listed in Table 2-2, through the HiBiT-qIP assay using GST protein fused with their monomeric form of the tags (Fig. 2-4). In these assays, the epitope-tagged GST proteins at seven concentrations, ranging from 0.825 ng (~0.025 nM) to 330 ng (~10 nM), were mixed with a fixed amount of cognate monoclonal antibody such that the binding curves reached a plateau. Preliminary IP experiments revealed that anti-IgG magnetic beads more efficiently captured monoclonal antibodies, irrespective of their IgG subclasses, than protein G magnetic beads (our unpublished data, also see Kimura *et al.*[47]). Thus, the IP reactions were performed using antibodies immobilised on anti-IgG magnetic beads in 1 mL of the stringent IP buffer (known as the typical RIPA buffer), which contains 0.1% SDS, 1% Triton X-100 and 0.1% sodium deoxycholate as the detergent in Tris-buffered saline (50 mM Tris-HCl [pH 7.5], 150 mM NaCl). This IP buffer composition was selected because similar conditions have often been used in standard IP[7,40,41] and ChIP experiments[48,49,50], and ChIP is currently one of the most important applications of IP. The antibody concentration used in the IP solution was empirically adjusted and varied from 20 pM to 0.2 nM depending on the affinity of the tested antibody/antigen pair (see Materials and Methods). Each K_d determination experiment was conducted in duplicate, and 14 data points were used for the curve-fitting analysis (Fig. 2-4). The error plots obtained from the K_d determination experiments showed a clearly defined minimum in the sum of squared residuals (SSR) (Fig. 2-4, right panels), validating the accuracy of the K_d value and the antibody concentration selected for each experiment.

A considerable variation in the K_d values was observed among the antibody clones examined, and these values ranged from 3.8×10^{-10} M for anti-HA (3F10) to 6.7×10^{-9} M for anti-HA (4B2) (Fig. 2-4A, B), but fell within a reasonable range of K_d values for high-affinity

monoclonal antibodies, which suggests that our method exhibits high validity. The comparison of the measured K_d values with those available from the literature revealed both similarities and differences (Table 2-3). The K_d value for anti-PA (NZ-1) against the dodecapeptide PA tag measured using our HiBiT-qIP assay was 7.1 x 10^{-10} M, which is close to the reported K_d value of 4.0 x 10^{-10} M obtained through a kinetic analysis using SPR[35]. The K_d value for anti-HA (4B2) obtained with our HiBiT-qIP assay was 6.3 x 10^{-9} M, which is not very different from the reported K_d value of 1.6 x 10^{-9} M obtained by the SPR method[35]. For anti-FLAG (M2), the obtained K_d value of 7.7 x 10^{-10} M deviated from those reported, which range from 3 x 10^{-9} M to 2.8 x 10^{-8} M[35,51,52]. This discrepancy might be due to differences in the position of the FLAG tag, the conditions used, including the buffer composition and pH, the method used, and/or the detection sensitivity[12,14].

Epitope tag antibody	K_d (nM)	Assay format/ method	References
FLAG (M2)	0.76	HiBiT-qIP	This study
	28	SPR	Fujii et al. 2014 [35]
	6.7	SPR	Wegner et al. 2002 [51]
	3	Saturation binding and Scatchard analysis	Firsov et al. 1996 [52]
HA (4B2)	6.6	HiBiT-qIP	This study
	1.6	SPR	Fujii et al. 2014 [35]
PA (NZ-1)	0.65	HiBiT-qIP	This study
	0.40	SPR	Fujii et al. 2014 [35]

Table 2-3. Comparison of the K_d values obtained in this study with those reported in the literature determined by other methods. To the best of our knowledge, the K_d values for the listed epitope tag antibodies are all the data available in the literature.

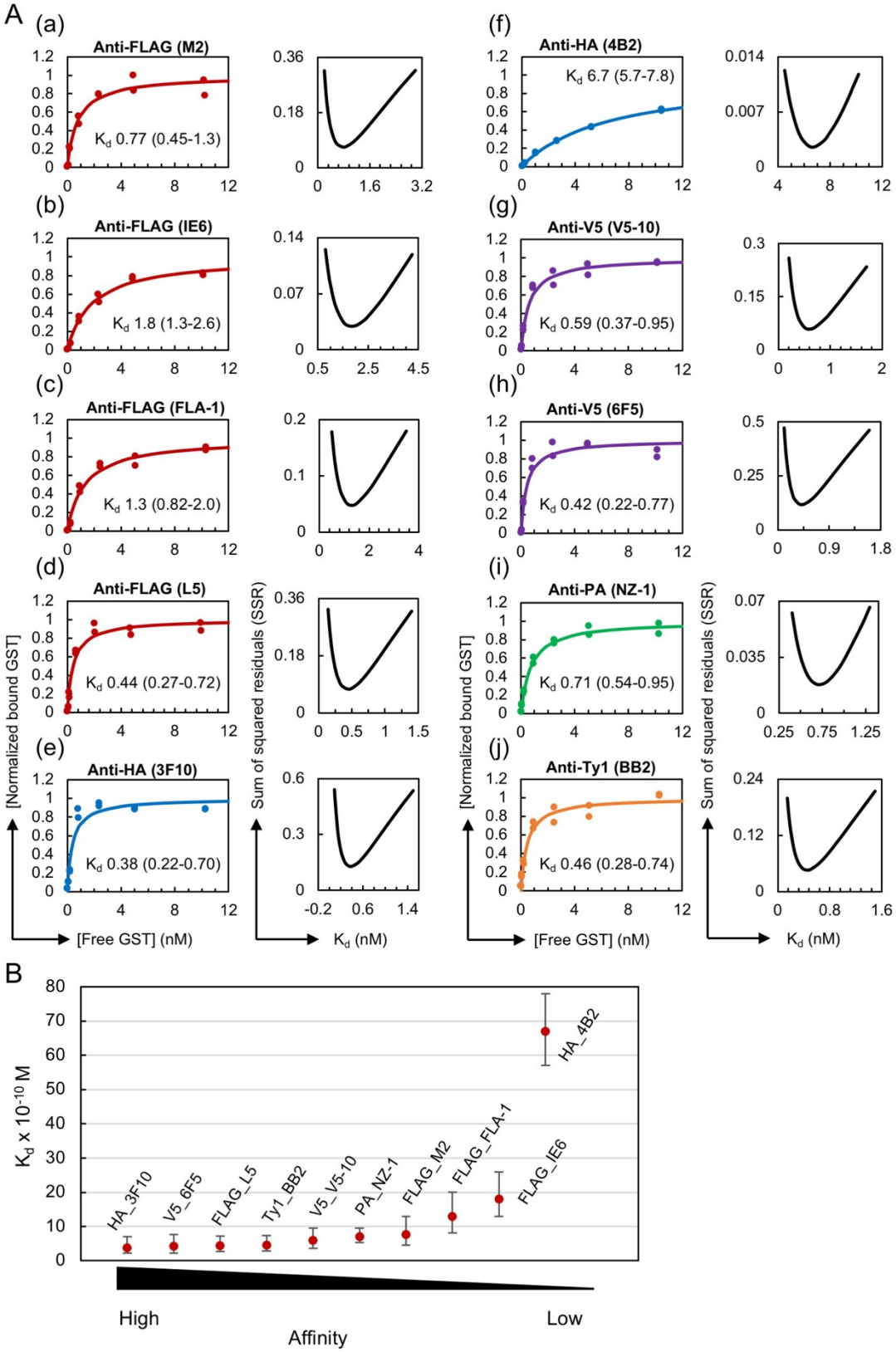


Figure 2-4. Considerable variation in K_d values among the epitope tag antibody clones. (**A**) (Left panel) Binding curves of the tested antibody clones against the monomeric form of the epitope tags. The antibody concentrations used for IP were as follows: 0.2 nM for anti-FLAG (M2, IE6, FLA-1 and L5) and anti-V5 (V5-10 and 6F5); 0.1 nM for anti-HA

(3F10 and 4B2) and anti-PA (NZ-1); and 0.05 nM for anti-Ty1 (BB2). (Right panel) Error curves for the best-fitting K_d. In each plot, the obtained apparent K_d value in nM is shown with the 95% confidence interval. (**B**) Affinity comparison of the antibody clones shown in panel A. Error bars depict the plus and minus confidence interval of the K_d value.

We repeated the measurements of the four antibody clones, anti-FLAG (M2), anti-HA (4B2), anti-PA (NZ-1) and anti-Ty1 (BB2), to assess the reproducibility of our HiBiT-qIP-based K_d determinations (Fig. 2-5Aa-d) and obtained a very similar apparent K_d value in all cases, which indicated that the developed method shows high reproducibility. Additionally, combined data plots were generated using the data from the two independent experiments shown in Figures 2-4 and 2-5, and these plots confirmed the reproducibility of the HiBiT-qIP assay (Fig. 2-5Ae-h).

Notably, the rat monoclonal anti-HA (3F10), anti-FLAG (L5) and anti-PA (NZ-1) antibodies displayed significantly lower apparent K_d values among the clones tested, suggesting the greater utility of rat monoclonal antibodies. Among the tested anti-FLAG antibody clones, anti-FLAG (L5) exhibited a considerably lower K_d value than the most widely used anti-FLAG (M2), consistent with the observation that the L5 clone detects FLAG-tagged proteins more efficiently than the M2 clone in Western blotting[53]. Interestingly, the anti-Ty1 (BB2) and anti-V5 (V5-10 and 6F5) antibody clones exhibited the highest affinity among the tested mouse clones, even though the Ty1 and V5 epitope tags have been less commonly used in IP experiments than the FLAG and HA tags. This finding suggests that the Ty1 and V5 epitope tags could perform similarly to or even better than the FLAG and HA tags in IP experiments, depending on the antibody used. These results together suggest the advantage of evaluating several different clones prior to performing IP experiments and thereby identifying the most suitable clone for each epitope tag that will be used in the experiments.

In the IP procedure described above, we used the antibody-bound anti-IgG beads to

capture the tagged GST proteins. Theoretically, this method measures the overall affinity of two interactions, namely, the epitope tag-antibody interaction and the antibody-anti-IgG bead interaction. In these IP reactions, however, excess amounts of anti-mouse or anti-rat IgG beads were used and most primary antibodies could be captured by the beads. Thus, it is very likely that our assay essentially measured the affinity of epitope tag-antibody interactions. To directly test this hypothesis, we used commercially available magnetic beads that have been covalently cross-linked to the anti-FLAG (IE6) mouse antibody or anti-PA (NZ-1) rat antibody. In both cases, we obtained K_d values that were slightly larger than those determined using the anti-IgG-bead-based protocol (Fig. 2-5B), which suggests that our assay actually measures the affinity of the epitope tag-antibody interaction.

Apparent K_d values could vary among different IP conditions, as noted in the Introduction. To examine these differences, if any, we performed an IP experiment using RIPA buffer without SDS because IP assays, particularly co-immunoprecipitation (Co-IP) assays, are often performed under relatively more native conditions. For this assay, the GST protein fused with a monomeric HA tag was prepared in native form and used with the anti-HA (3F10) antibody. The assay yielded a K_d value that was comparable to that obtained with SDS-containing RIPA buffer (Fig. 2-5C), which indicated that anti-HA (3F10) performs equally well under these two conditions.

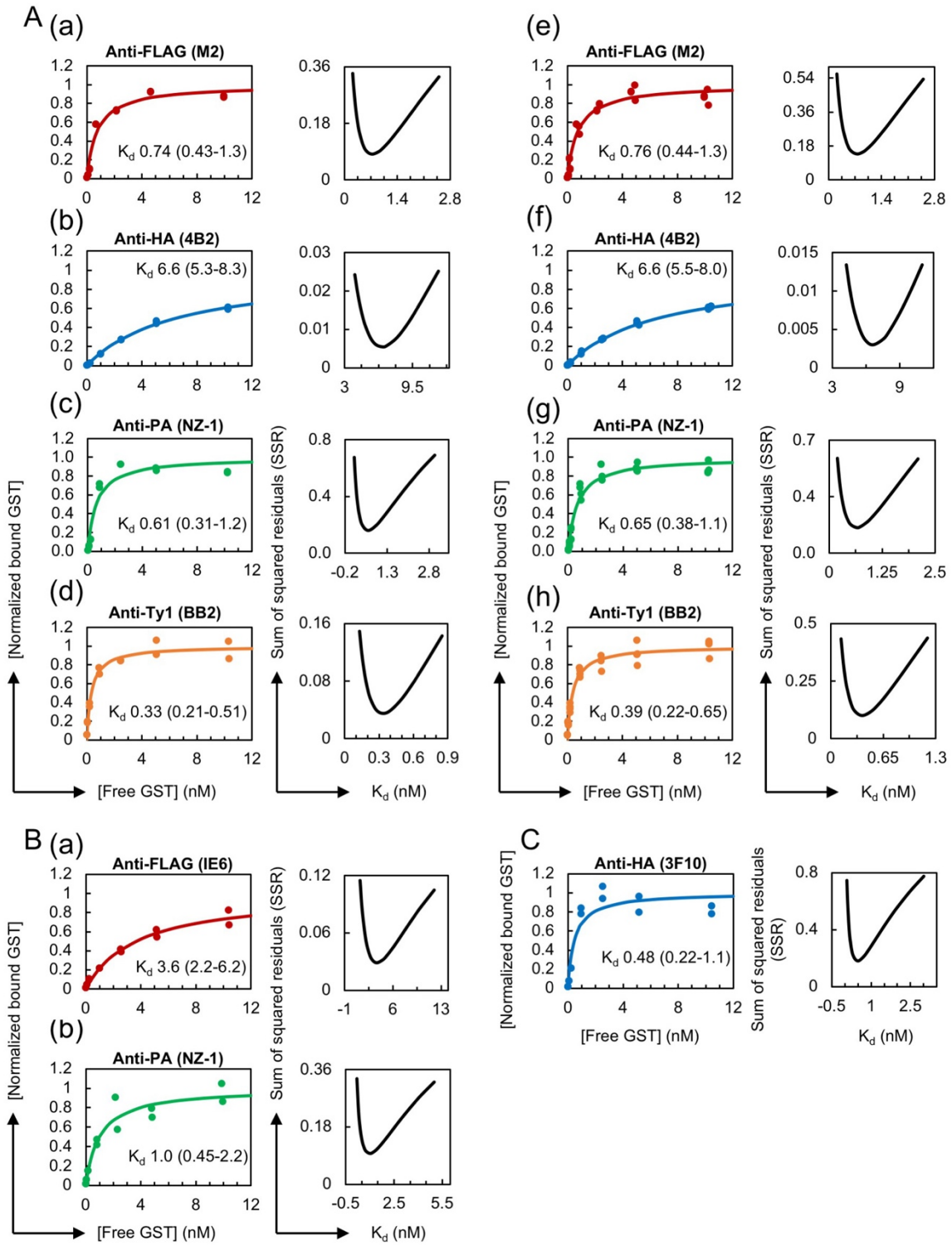


Figure 2-5. Validity and reproducibility of the HiBiT-qIP assay. (**A**) Reproducibility of the HiBiT-qIP-based K_d determination. (**a-d**) K_d determination experiments were repeated for four monoclonal antibody clones: anti-FLAG (M2), anti-HA (4B2), anti-PA (NZ-1) and anti-Ty1 (BB2). (Left panel) Binding curves of the antibody clones tested against the monomeric form of the epitope tags. The antibody concentrations used for IP were as follows: 0.2 nM for anti-FLAG (M2) and anti-HA (4B2); 0.1 nM for anti-PA (NZ-1); and

0.05 nM for anti-Ty1 (BB2). (**e-g**) Binding curves plotted with data obtained from two independent experiments, shown in Fig. 2A and Fig. 3A. (**B**) IP performed using magnetic beads covalently cross-linked to anti-FLAG and anti-PA antibodies provided comparable K_d values. (Left panel) Binding curves of the antibody clones tested against the monomeric form of the epitope tags. The concentrations of anti-tag antibodies attached to the beads in IP were as follows: 1 nM for anti-FLAG (IE6) and 0.2 nM for anti-PA (NZ-1). (**C**) IP performed under native conditions using RIPA buffer without SDS provided a comparable K_d value. (Left panel) Binding curve of the anti-HA (3F10) clone against the monomeric form of HA. The concentration of the antibody used for IP was 0.1 nM. (**A,B,C**) (Right panel) Error curves for the best fit K_d. In each plot, the obtained apparent K_d value is shown with the 95% confidence interval.

2-2-4 A significant increase in affinity was observed with the use of epitope tags in dimeric or trimeric form.

The dimeric or trimeric form of epitope tags has frequently been used in a variety of immunoassays to enhance their sensitivity[38,54–56], but the effects of multimerisation in immunoprecipitation have not yet been quantitatively characterised. To address this problem, we measured the apparent K_d values for the dimeric and trimeric forms of the epitope tags based on the assumption that a one-to-one interaction primarily occurs between the antibody and the multimerised epitope tag peptide under our assay conditions (see Discussion). Here, we thus use the term "apparent K_d" as an artificial parameter to describe the interaction between the antibody and the multimerised tag by considering the multimerised tag as a single binding site. We produced GST proteins with trimeric forms of FLAG and HA and GST proteins with dimeric and trimeric forms of V5, PA and Ty1 (Table 2-1, Fig. 2-2). Our trimeric form of FLAG consisted of simple direct repeats of DYKDDDDK, and was thus not identical to the original 3xFLAG sequence DYKDHDGDYKDHDIDYKDDDDK, in which modified FLAG sequences are used in the first and second positions[56]. We selected this simple repeated form because the original 3xFLAG was optimised for the traditional anti-FLAG (M2) clone and might thus not be recognised by the newly developed anti-FLAG monoclonal antibodies used in this study. The use of these epitope-tagged GST proteins in the HiBiT-qIP assay revealed a several-fold increase in the apparent affinity compared with

that obtained with the monomeric forms (Figs. 2-6 and 2-7, Table 2-2). The comparison of the mono-, di- and trimeric forms showed a gradual increase in affinity depending on the number of epitope tags (Fig. 2-7 C-E), indicating a clear positive correlation between the apparent affinity and the number of epitopes; however, the differences in affinity between the dimeric and trimeric forms were rather small, particularly those of anti-V5 (6F5) (Fig. 2-7Cb).

A comparison of the mono- and trimeric forms of FLAG and HA revealed a considerable increase in the affinity of all the tested antibodies with the use of their trimeric forms. The anti-HA 3F10 clone with the lowest K_d value in our comparisons exhibited increased affinity against the trimeric form of HA, and showed the highest affinity among the clones tested (Figs. 2-4B, 2-6Ba, 2-7Ba). Interestingly, the difference in apparent affinities among the four anti-FLAG monoclonal clones decreased when examined against the trimeric form of the FLAG tag. This phenomenon was also clearly observed with anti-HA (4B2).

Taken together, these results indicate that the dimerisation and trimerisation of epitope tags clearly increase the apparent affinity of antibodies under IP conditions. In addition, the results suggest that in cases in which high-affinity antibodies are unavailable, low-affinity antibodies might be successfully used in IP experiments if combined with multimeric forms of the epitope tags.

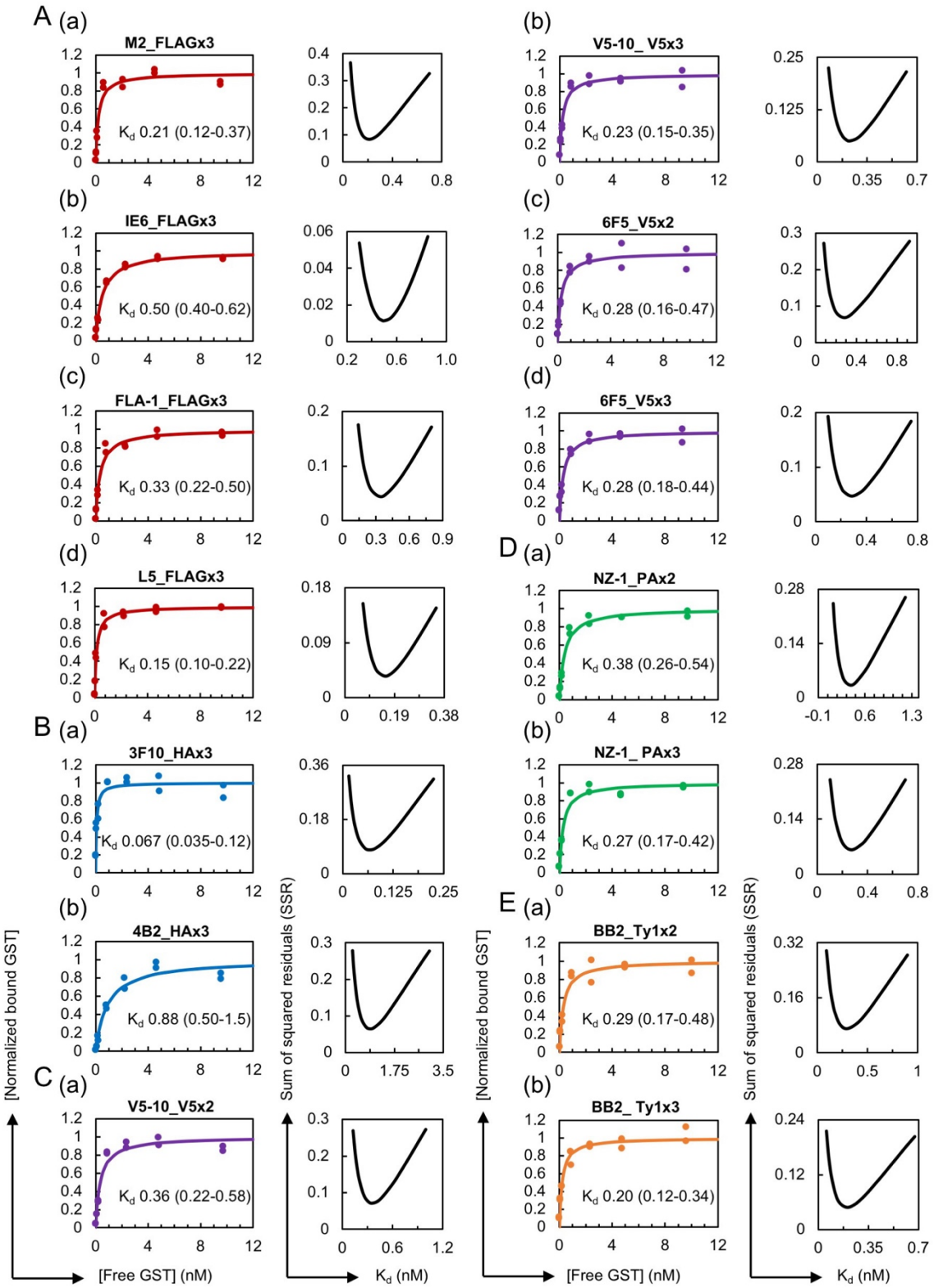


Figure 2-6. Increase in the antibody affinity against the dimeric or trimeric form of the epitope tags. (Left panel) (**A**) Binding curves of anti-FLAG M2 (a), IE6 (b), FLA-1 (c) and L5 (d) clones against FLAGx3. The antibody concentrations used for IP were 0.1 nM for (a-c) and 0.05 nM for (d). (**B**) Binding curves of the anti-HA 3F10 (a) and 4B2 (b) clones

against HAx3. The antibody concentrations used for 3F10 and 4B2 were 0.02 nM and 0.1 nM, respectively. (**C**) Binding curves of the anti-V5, V5-10 and 6F5 clones against V5x2 (a,c) and V5x3 (b,d). The antibody concentration used for V5-10 was 0.05 nM for both forms. The concentrations of the antibodies used for 6F5 were 0.1 nM for V5x2 and 0.05 nM for V5x3. (**D**) Binding curves of the anti-PA NZ-1 clone against PAx2 (a) and PAx3 (b). The concentrations of the antibody used for IP were 0.1 nM for PAx2 and 0.05 nM for PAx3. (**E**) Binding curves of the anti-Ty1 BB2 clone against Ty1x2 (a) and Ty1x3 (b). The concentration of the antibody used for IP was 0.05 nM for both forms. (Right panel) Error curves for the best-fitting K_d. In each plot, the obtained apparent K_d value in nM is shown with the 95% confidence interval.

2-2-5 Tag multimerisation greatly improved the efficiency of IP from crude cell lysates

Because a significant increase in affinity was observed with the use of epitope tags in multimeric form, we questioned the resulting effects on the efficiency of IP from crude cell lysates, which is closer to real experimental conditions. To answer this question, we synthesised mRNAs encoding the zebrafish transcription factor Sox3 tagged with a monomeric or trimeric form of the FLAG tag and HiBiT, expressed these proteins in zebrafish embryos at near-endogenous levels, and prepared crude cell lysates in RIPA buffer containing SDS. We immunoprecipitated FLAG-tagged Sox3 using an anti-FLAG (IE6) antibody and quantified the amount of recovered Sox3 proteins. Specifically, Western blotting with an anti-Sox3 antibody was used to determine the relative amounts of Sox3, and HiBiT blotting using GST-FLAGx3-HiBiT as a standard was performed to convert the relative amounts to absolute amounts (Fig. 2-8, Fig. 2-9). In accordance with the differences in affinity, a considerable improvement in IP recovery was clearly observed with the use of the trimeric form of the FLAG tag (Fig. 2-8A,C), and this effect was more pronounced if a limited amount of antibody was used. Interestingly, the comparison of these observed recovery rates with those calculated theoretically based on the K_d value revealed a substantial difference, particularly for the monomeric FLAG tag (Fig. 2-8C, see Discussion).

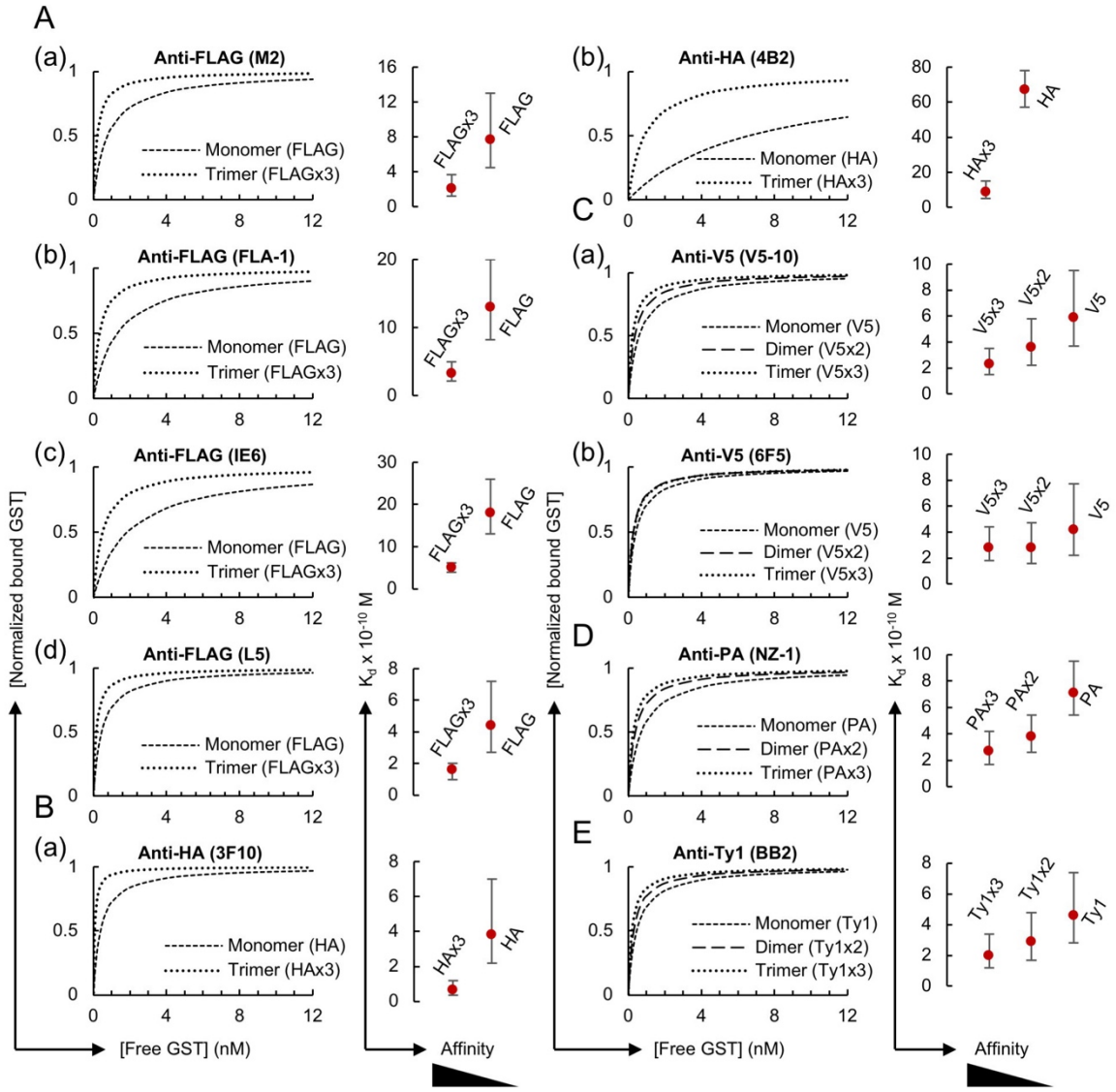
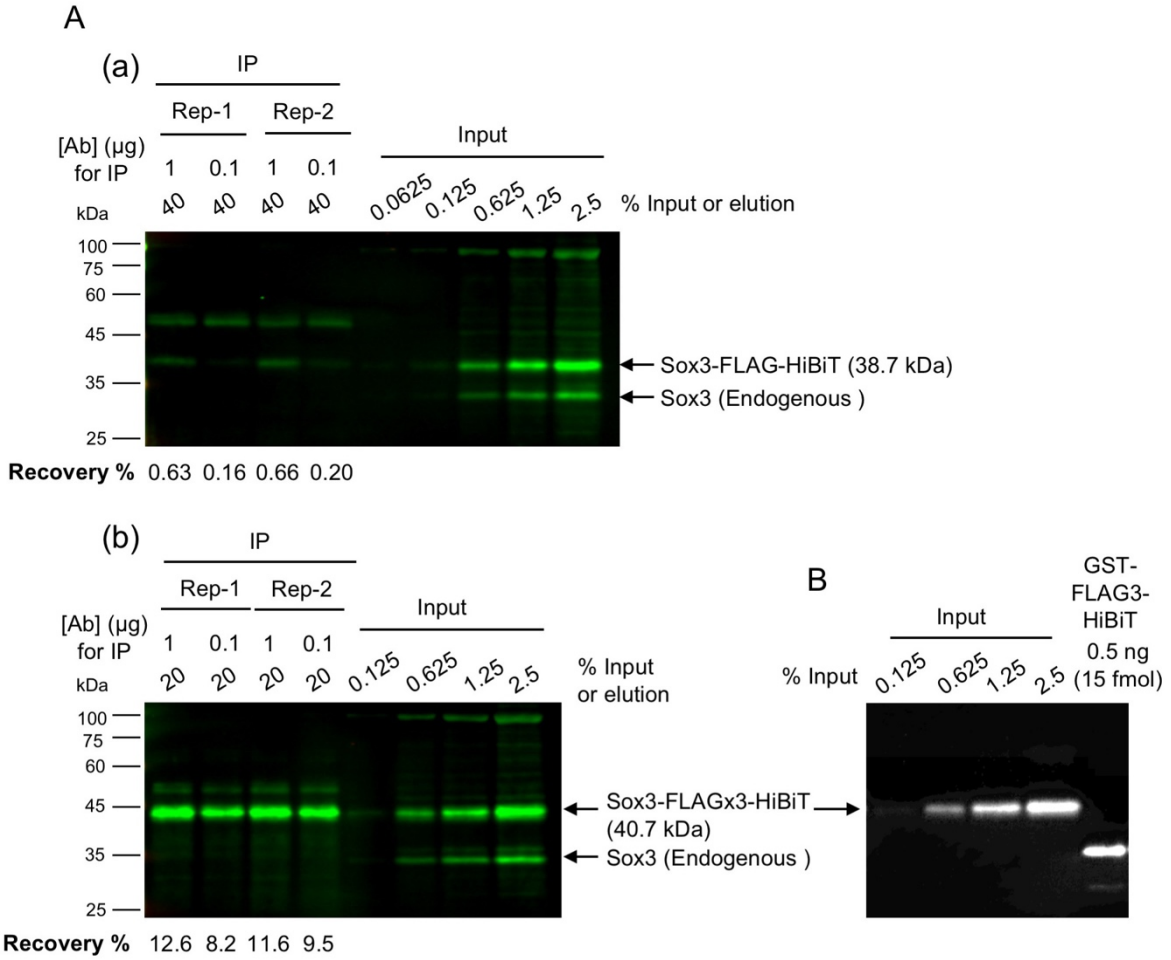


Figure 2-7. Effect of tag multimerisation on the apparent affinity of the antibodies. (Left panel) The binding curves of anti-epitope tag antibody clones against monomeric, dimeric and/or trimeric forms of the epitope tags shown Figures 2-4 and 2-6 are simultaneously plotted for comparison purposes. (Right panel) Affinity comparison. The error bars depict the 95% confidence intervals for the K_d values.



C

	Antibody used for IP		Input protein		Bound protein (Average)		Recovery percentage (Average)	Theoretical recovery		
	µg	nM	ng	nM	ng	pM		ng	nM	Percentage
Sox3-FLAG-HiBiT	1	6.67	18.9	0.487	0.12	3.1	0.65	16.64	0.43	88.4
	0.1	0.67	18.9	0.487	0.03	0.9	0.18	7.74	0.20	40.2
Sox3-FLAGx3-HiBiT	1	6.67	27.6	0.677	3.34	82.0	12.12	26.46	0.65	96.2
	0.1	0.67	27.6	0.677	2.43	59.7	8.81	17.91	0.44	64.4

Figure 2-8. Effect of tag multimerisation on the efficiency of IP from crude cell lysates. (A) Infrared fluorescent Western blot detection of the FLAG-tagged Sox3 proteins. The Sox3 protein tagged with FLAG either in monomeric (a) or trimeric (b) form in IP samples and the dilution series of input lysates were detected using anti-Sox3 antibody. IP was performed in duplicate and under two IP conditions, i.e., 1 µg or 0.1 µg of anti-FLAG (IE6) antibody was incubated with the embryo lysates in 1 mL of IP solution. The recovery rates (in percentages) of FLAG-tagged Sox3 are shown at the bottom of each blot. The two blots in panels (a) and (b) were processed in parallel and scanned simultaneously. (B) HiBiT detection of input dilutions of FLAGx3-tagged Sox3 on the same membrane shown in (Ab) and 0.5 ng (15

fmol) of FLAGx3-tagged GST, which was used as the standard for absolute protein quantification. (C) Comparison of the experimental and theoretical IP recovery values. The averaged recovery rates of duplicated IP experiments and theoretical recovery rates are shown. The process used for the theoretical recovery calculation is given in the Materials and Methods section.

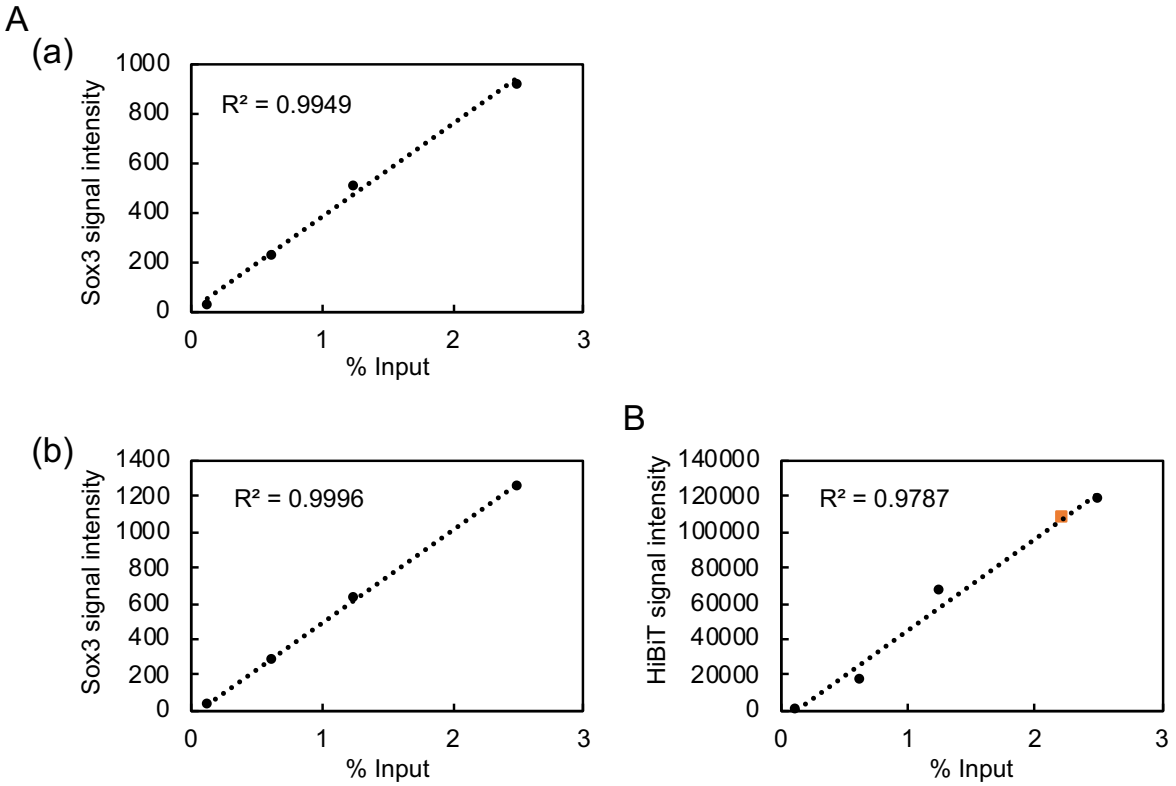


Figure 2-9. Standard curves demonstrating linear correlation of signal intensity with the %input. (A) Near infrared fluorescence (NIR) signals were linear across the entire range of %input tested, with $R^2 > 0.99$ for both the Sox3 protein tagged with FLAG either in (a) monomeric or (b) trimeric form. (B) HiBiT derived chemiluminescence was linear across the entire range of %input tested, with $R^2 > 0.97$ for FLAGx3-tagged-Sox3 protein. The orange square mark the related signal intensity and %input for the 0.5 ng of FLAGx3-tagged-GST protein.

2-3 Discussion

2-3-1 Determination of K_d values through the HiBiT-qIP assay.

In the current study, we employed the NanoLuc-based HiBiT system[20,21] to establish our HiBiT-qIP assay for determining the K_d values for protein-protein interactions in solution. We applied this method to measure the K_d values for interactions between certain epitope tags and a series of their cognate monoclonal antibodies under the IP conditions generally used in ChIP. We assumed that the antigen-antibody interaction reaches equilibrium during the overnight IP and that the steady-state dissociation constant could be determined by measuring the amount of immunoprecipitated proteins after a brief wash. Due to the high sensitivity of the HiBiT solution assay, in which a sub-picogram protein amount can be detected within the linear response region, we were able to perform titration experiments with a wide range of antigen concentrations. In fact, the apparent K_d values for the tested monoclonal antibody clones were found to be between 10^{-9} M and 10^{-11} M, which is the typical range of the reported K_d values for high-affinity antigen-antibody interactions[7,45]. Furthermore, for the anti-PA antibody, we obtained a K_d value similar to that reported[35]. To evaluate the accuracy of the K_d values obtained and the antibody concentration selected for each experiment, the 95% confidence intervals (CIs) were assessed as previously described[57,58]. In most cases, we were able to obtain K_d values with 95% confidence intervals from half of the K_d to twice the K_d. These findings strongly suggest that the HiBiT-qIP assay is able to measure K_d values with reasonably good accuracy.

Notably, although we used the HiBiT-qIP assay to measure the K_d values for anti-epitope tag monoclonal antibodies, this method could be used to obtain the K_d values for any type of monoclonal antibody as long as the target protein can be tagged with HiBiT. Moreover, in theory, the K_d values for any type of protein-protein interaction in solution might be obtained if one can prepare an appropriate combination of a HiBiT-tagged protein

and another protein that can be captured with beads through covalent crosslinking or a high affinity interaction such as avidin-biotin.

Only a few methods have been developed for quantitatively characterising antigen-antibody interactions under IP conditions, which has resulted in difficulties in selecting suitable antibodies for demanding IP applications such as ChIP[2,6]. To overcome this problem, a quantitative peptide immunoprecipitation (peptide IP) assay in a ChIP-like format was developed by Nishikori *et al.*[59]. In their method, a biotinylated antigen peptide is incubated with antibody-bound protein A (or G) polystyrene beads in solution. The captured peptide is then linked to fluorescently labelled streptavidin and quantified using flow cytometry. The advantage of their assay is that it is readily applicable if a biotinylated antigen peptide is available, but an inherent drawback is that the antigen-antibody complex might dissociate during flow cytometry and the IP wash process, which could lead to underestimation of the antibody affinity. Our HiBiT-qIP assay has the same dissociation problem during the IP wash process, but the effect can be minimised by performing the wash process reasonably rapidly. Our HiBiT-qIP assay and the peptide IP assay developed by Nishikori *et al.*[59] technically measure the overall affinity of all the interactions involved in the process: the former assay measures the overall affinity of the antibody-antigen and antibody-bead interactions, whereas the latter measures the overall affinity of the antibody-peptide, antibody-bead and biotin-streptavidin interactions. However, as discussed by Hattori *et al.*[60], these IP-based assays appear to essentially measure the affinity of the antibody-antigen interactions because most of the primary antibodies are expected to be captured by the capture beads, which are added in excess amounts in IP reactions. Consistent with this notion, we obtained similar apparent K_d values even with magnetic beads that had been covalently cross-linked to tag antibodies (Fig. 2-5B). However, this finding also implies that it might be difficult to measure interactions with affinities higher than those of the

antibody-bead interaction in these standard IP-based assays.

At present, the most widely used methods for determining the K_d values of an antigen-antibody interactions are SPR technology[10,11] and KinExA[12,13]. KinExA allows the direct measurement of the equilibrium binding affinity of interactions between molecules in solution after an equilibrium is reached. In contrast, in the SPR approach, the kinetic parameters, the association rate constant (k_{on}) and the dissociation rate constant (k_{off}) are first determined, and these parameters are then used to calculate K_d as $K_d = k_{off}/k_{on}$. Due to methodological similarities, a K_d value obtained using the HiBiT-qIP assay might be closer to that measured by KinExA. Because SPR technology might not accurately measure very slow k_{off} values in a standard flow mode, it might be difficult to measure high-affinity interactions with low k_{off} values using this method[61]. In contrast, the HiBiT-qIP assay cannot be applied to measure interactions with high k_{off} values because this method involves a wash process after IP, as discussed above. The advantages of the HiBiT-qIP assay compared with SPR and KinExA are summarised as follows: (1) this method can directly measure apparent K_d values under specific IP buffer conditions, (2) this method can be performed with small amounts of antibody and antigen, and (3) this method only uses a standard luminometer and thus provides a more accessible approach for determining K_d values.

2-3-2 Evaluation of epitope tag antibodies.

It has been suggested that immunoprecipitation experiments normally require antibody affinities of at least 10^{-8} M for efficient recovery because they rely on the formation of an antigen-antibody complex in solution at relatively low concentrations of the antigen[7]. The monoclonal anti-epitope tag antibodies tested in this study all meet this criterion, which is consistent with the fact that they are supplied as antibodies that can be used for IP. For more demanding IP experiments, however, higher-affinity antibodies with K_d values equal to or

less than 10^{-9} M might be required.

Our results clearly show a several-fold increase in the apparent affinity by the use of any of the epitope tags in dimeric or trimeric form. This effect can be explained in part by the fact that the use of multimeric tags leads to a simple increase in molar concentration of each monomeric tag. Additionally, the close proximity of the epitopes on the multimeric tags might facilitate rebinding of the antibody to the neighbouring sites. On the other hand, bivalent binding of an antibody to a single multimerised tag may not be possible due to a large distance between the two antigen binding sites relative to the length of tag polypeptides[62]. Moreover, simultaneous binding of a multimeric tag to neighbouring antibodies on a magnetic bead also seems to be a rare case because surface density of the antibodies on the bead is calculated to be low (less than one antibody/2300 nm^2) under our IP conditions. Historically, Hernan *et al.*[56] discovered that the Western blot detection limit can be improved by more than 10-fold by tagging a target protein with a sequence containing two additional FLAG epitopes in tandem with the original FLAG sequence (3x FLAG). Since then, the 3x FLAG epitope tag has been widely used due to its enhanced sensitivity in affinity isolation and immunohistochemical detection. Remarkably, epitope tagging with the triple-FLAG tag facilitated the IP of low-abundance proteins at near-endogenous levels, whereas the FLAG monomer failed to immunoprecipitate the proteins[55]. Consistent with this, we observed a considerable improvement in IP recovery with the use of the trimeric FLAG tag (Fig. 2-8). This enhancing effect can be clearly observed in IP experiments performed with limited amounts of the target protein. In fact, Zhang *et al.*[63] observed equal precipitation of both the monomeric and trimeric forms of FLAG-tagged target proteins in their IP experiments. Consistent with this, under our IP conditions, we clearly observed a substantial increase in the HiBiT-derived signal from the immunoprecipitate of the FLAG trimer compared with that obtained with the FLAG monomer only if lower amounts of

FLAG-tagged GST were used (Fig. 2-7). The IP of FLAG-tagged Sox3 from the crude cell lysates with the monomeric FLAG tag exhibited a substantially lower IP recovery compared with that obtained with the trimeric FLAG tag. This difference cannot be explained by the affinity difference. However, it is possible that some protein components in the crude lysate might strongly inhibit the antibody-tag interaction and this inhibition might have been overcome by tag multimerisation. Interestingly, our results show that not only the FLAG tag but also the rest of the epitope tags we tested exhibited improved affinity and therefore an increased IP yield when used in multiple tandem repeats. This finding reflects the wide use of epitope tags in their multimerised forms, although the biochemical basis has been rarely examined.

The utility of an antibody in IP is critically dependent on its specificity in addition to its affinity, although we did not address this point in this study. A recent study by Macron *et al.*[64] showed that antibody selectivity and specificity in IP can be effectively characterised by quantifying the abundance of all the proteins in the immunoprecipitates. This approach is complementary to that presented in this paper, and these two approaches can support each other. Thus, high affinity is not the sole criterion for selecting good antibodies for IP experiments but might be the most important factor because a high-affinity interaction enables the performance of IP experiments under stringent conditions, which would result in an increase in specificity. Overall, this comparison of monoclonal antibodies against the commonly used epitope tags will prove to be a valuable resource for future IP-related experiments.

Finally, we suggest that our HiBiT-qIP assay might also be useful in quantitative monitoring of IP experiments. As mentioned in the Introduction and elaborated in Hakhverdyan *et al.*[65], the efficiency of IP is strongly influenced by reagents used in IP solution such as salts, buffers and detergents. The efficiency of IP is also affected by the

complexity of protein samples, as observed in this study (Fig. 2-8). However, it is unpredictable how these factors affect the performance of IP[65]. Thus, it would be necessary to explore the parameters affecting the efficiency of IP, particularly when the target protein is expressed at near-endogenous levels. Under such circumstances, HiBiT-qIP could facilitate the evaluation of various IP parameters through quantitative analysis of the immunoprecipitated proteins tagged with HiBiT.

2-4 Materials and Methods

2-4-1 Plasmid DNA construction for epitope-tagged GST protein expression.

Using pGEX-6P-1 (GE Healthcare) as the parental vector, the coding region of the glutathione S-transferase (GST) gene was fused in-frame to a series of composite tags at the *Xho*I site in the multi-cloning sites. Each of the composite tags contained one of the epitope tags, namely, FLAG, HA, PA, V5 and Ty1 in either monomeric, dimeric or trimeric form, followed by a TEV protease cleavage site, a biotin acceptor domain (Bio tag) and, most C-terminally, the HiBiT epitope tag. The exact nucleotide and amino acid sequences of the composite epitope tags are listed in Figure 2-2.

2-4-2 Epitope tag antibodies.

The monoclonal antibodies used for immunoprecipitation are provided, along with suppliers, clone IDs, host species, IgG subclasses and the type of magnetic beads, in Table 2-2. The product numbers of these antibodies are as follows: anti-FLAG (M2), F1804; anti-FLAG (FLA-1), M185-3S; anti-FLAG (IE6), 018-22386; anti-FLAG (L5), 637301; anti-HA (3F10), 11867423001; anti-HA (4B2), 010-21883; anti-PA (NZ-1), 016-25861; anti-V5 (V5-10), V8012; anti-V5 (6F5), 017-23593; and anti-Ty1 (BB2), SAB4800032.

2-4-3 Expression and purification of the epitope-tagged GST proteins.

The GEX-6P-1-derived plasmids containing the various composite tag sequences were transformed into *Escherichia coli* (JM109) cells for protein expression. A single colony of transformed *E.coli* cells was inoculated in 2xYT medium and incubated overnight at 37 ºC with vigorous shaking. The culture was diluted 1:100 into 5 mL of fresh 2xYT medium and incubated at 28 ºC with shaking until the A_{600} reached 0.6-0.8. Protein expression was induced by the addition of IPTG (0.1 mM), and the cells were then incubated for an

additional 3 hours, pelleted by centrifugation, resuspended in TBS (Tris-buffered saline: 20 mM Tris-HCl [pH 7.5] and 150 mM NaCl) and lysed by sonication (Bioruptor, Cosmo Bio) until the suspension became clear. The soluble fraction, which was separated from the insoluble fraction by centrifugation (15,000 x g, 4 °C, 5 min), was mixed with glutathione–Sepharose beads (GE Healthcare) and rotated for 10 min at room temperature (RT). The resin was collected by brief centrifugation, and unbound proteins were washed away from the beads with TBS. The GST fusion proteins were then eluted by the addition of nuclei lysis buffer (50 mM Tris-HCl [pH 7.5], and10 mM EDTA, 1% [w/v] SDS). In addition, to elute the GST proteins in native form, glutathione elution buffer consisting of 10 mM reduced glutathione and 50 mM Tris-HCl (pH 8.0) was used. The purified proteins were subsequently separated using 10% SDS-polyacrylamide gels, stained with Coomassie Brilliant Blue G-250 and scanned with an Odyssey instrument (LI-COR) using the 700-nm channel[39]. The scanned image was used to quantify the protein amounts using BSA as the standard.

2-4-4 Nano-Glo HiBiT blotting.

To detect the HiBiT-derived signal from the GST proteins fused with monomeric, dimeric and trimeric forms of the epitope tags and the HiBiT peptide, equal amounts of the purified proteins were separated on 10% SDS-polyacrylamide gels and transferred to nitrocellulose membranes. The protein-transferred membranes were incubated in TBST for 30 min, and this medium was then replaced with Nano-Glo HiBiT blotting reagent containing LgBiT protein (Promega). After 1 hour of incubation at RT, the substrate furimazine was added, and the incubation was continued for another 5 min. The blot was imaged using a chemiluminescence imager with a CCD camera (Fusion, Vilber Lourmat).

2-4-5 Immunoprecipitation.

Varying concentrations of the purified epitope-tagged GST proteins were prepared in nuclei lysis buffer and diluted 10-fold with ChIP dilution buffer (50 mM Tris-HCl [pH 7.5], 167 mM NaCl, 1.1% Triton X-100, and 0.11% [w/v] sodium deoxycholate) to obtain the RIPA buffer composition (50 mM Tris-HCl [pH 7.5], 150 mM NaCl, 1.0% Triton X-100, 0.1% sodium deoxycholate, and 0.1% SDS). One millilitre of the diluted protein sample was mixed with 15 µL of anti-mouse or anti-rat IgG magnetic beads (~10 mg/mL, Invitrogen) in PBS supplemented with 0.5% BSA (Rockland Immunochemicals) that had been pre-bound to 30 ng (or lower, i.e., 15 ng, 7.5 ng or 3 ng, as stated in the figure legends) of the anti-epitope tag antibody. The beads were incubated overnight at 4 ºC and then washed twice with ice-cold HEPES-RIPA buffer (50 mM HEPES [pH 7.6], 1 mM EDTA, 0.7% [w/v] sodium deoxycholate, 1% Triton X-100, and 0.5 M LiCl) and once with ice-cold TBS buffer. Any residual TBS was removed by centrifugation followed by aspiration. The immunoprecipitated protein was then eluted with 20 µL of elution buffer containing 0.1% (w/v) SDS, 50 mM Tris-HCl, 10 mM EDTA and 25 mM DDT for 5 min at 95 ºC. The supernatant was collected through a brief centrifugation (18,800 x g, 1 min). For each K_d determination, a fixed amount of antibody was used, while the epitope-tagged GST proteins were titrated as a dilution series ranging from 0.825 ng/mL (~0.025 nM) to 330 ng/mL (~10 nM) to obtain seven data points. Each point was analysed in duplicate independent samples to ensure the determination of an accurate K_d. The amount of antibody used was empirically optimised based on preliminary experiments such that the antibody concentration was maintained at a level lower than the K_d value. Specifically, the antibody amount that maintained at least half of the epitope-tagged GST protein in the unbound state, particularly for the data point corresponding to the lowest GST concentration, was considered the best suitable antibody concentration for IP. Given that the molecular mass of IgG is 150 kDa, 30 ng, 15 ng, 7.5 ng and 3 ng of an antibody in 1 mL

of immunoprecipitation reaction buffer correspond to concentrations of 0.2 nM, 0.1 nM, 0.05 nM and 0.02 nM, respectively. In the experiment involving magnetic beads covalently cross-linked to tag antibodies, we used anti-DYKDDDDK tag antibody magnetic beads (clone IE6, Wako) and MagCapture HP anti-PA tag antibody magnetic beads (clone NZ-1, Wako).

2-4-6 HiBiT detection assays.

The immunoprecipitated samples were diluted 100-fold using PBS containing 0.01% BSA and 0.1% Triton X-100, and 20 μL of these diluted samples was mixed with an equal volume of Nano-Glo HiBiT Lytic Reagent (Promega), consisting of Nano-Glo HiBiT Lytic Buffer, Nano-Glo HiBiT Lytic Substrate and LgBiT protein. This mixture was incubated for 10 min at RT, and the luminescence was measured using a Mithras LB940 plate reader (Berthold Technologies) with an integration time of 1 s. The amounts of the HiBiT tag were calculated using the same epitope-tagged GST protein as the standard.

2-4-7 Determination of apparent K_d.

The overnight incubation of IP samples at 4 ºC is expected to allow the binding reaction between the antigen and antibody to reach equilibrium. The bound epitope-tagged GST proteins were eluted, and the amount was determined using the HiBiT detection assays as described above. The apparent K_d was determined by fitting the data to the following equation[66]:

$$[L_b]/[L_{b\_max}]= [L_f]/(K_d+[L_f]),$$

where $[L_b]$ is the bound GST concentration (observed), $[L_{b\_max}]$ is the maximum bound GST concentration (calculated), and $[L_f]$ is the free GST concentration ($[L_{total}]$-$[L_b]$).

Nonlinear least-squares data fitting was accomplished using the Solver add-in regression tool

built in Microsoft Excel. Here, we first obtained the value of [L_{b_max}], and using these values, we then re-plotted the data to draw the final fitted curves that are shown in the figures, in which [L_b]/[L_{b_max}] is the normalised bound GST value. To assess the best-fit parameter values returned by the nonlinear regression, a 95% confidence interval was calculated using Fisher's F distribution, as elaborated by Kemmer *et al.*[57].

2-4-8 mRNA synthesis and zebrafish embryo microinjection.

To construct plasmids for the synthesis of mRNA encoding the zebrafish Sox3 protein tagged with a monomeric or trimeric form of the FLAG tag and the HiBiT tag, we inserted the zebrafish *sox3* coding sequence and the composite epitope tags into pCS2. The capped mRNAs for these FLAG-tagged Sox3 proteins were transcribed in vitro from linearised vectors using the mMessage mMachine SP6 kit (Ambion, ThermoFisher). Zebrafish embryos were obtained from the natural mating of wild-type TL fish and reared at 28.5 ºC in 0.03% Red Sea salt solution. Approximately 1 nL of solution containing FLAG-tagged Sox3 mRNA at a concentration of 10 ng/μL was injected into one -cell-stage embryos. The mRNA encoding the Venus fluorescent protein was included in the injection solution at a concentration of 50 ng/μL and used as a reporter to confirm the success of the microinjection. All zebrafish experiments were conducted in accordance with the Fundamental Guidelines for Proper Conduct of Animal Experiment and Related Activities in Academic Research Institutions under the jurisdiction of the Ministry of Education, Culture, Sports, Science and Technology of Japan using protocols approved by the Animal Experiments Committee of Osaka University. The same protocols were used at Kochi University of Technology.

2-4-9 IP followed by Western blotting using zebrafish embryo lysates.

Microinjected zebrafish embryos at the 70–80% epiboly stage were enzymatically

dechorionated with Pronase (2 mg/mL) and deyolked as described by Link *et al.*[67] One embryo/µL of embryo sample was prepared in nuclei lysis buffer containing the complete protease inhibitor cocktail (Roche). For each IP experiment, an embryo lysate equivalent to 100 embryos was sonicated and diluted 10-fold with ChIP dilution buffer containing protease inhibitors. IP was performed as described above, and proteins were eluted in 100 µL of SDS sample buffer containing 50 mM DTT. The samples were denatured by exposure to heat at 95 ºC for 10 min, and after a brief centrifugation (18,800 x g, 1 min), the supernatant was collected. The immunoprecipitated samples were then separated by SDS-PAGE along with a dilution series of input samples and transferred to nitrocellulose membranes (BioTrace NT, Pall Corporation) using a TransBlot Cell (Bio-Rad). The membrane was then blocked with Odyssey blocking buffer (1:1 diluted with TBS), rinsed with TBST buffer (TBS and 0.05% Tween-20), and probed with anti-Sox3 (GTX132494, GeneTex) at 0.5 µg/mL diluted in Can Get Signal solution 1 (Toyobo). The Sox-3 antibody was detected with goat anti-rabbit IgG-IRDye800 using an Odyssey CLx infrared imaging system (LI-COR Biosciences). The band intensities of Western blot images were quantified using Image Studio software (LI-COR Biosciences).

2-4-10 Theoretical IP recovery.

The K_d values obtained for the anti-FLAG (IE6) against the monomeric and trimeric forms of FLAG tag were used to calculate the theoretical IP recovery rates using the following equation:

$b = [(K_d+L+R)-\sqrt{(K_d+L+R)^2-4LR}]/2$, where b is the concentration of bound antibody, L is the total concentration of antigen, and R is the total concentration of binding sites on the antibody[66].

2-5 References

1. Bordeaux, J. *et al.* Antibody validation. *Biotechniques* **48,** 197–209 (2010).
2. Egelhofer, T. A. *et al.* An assessment of histone-modification antibody quality. *Nat. Struct. Mol. Biol.* **18,** 91–93 (2011).
3. Weller, M. G. Ten basic rules of antibody validation. *Anal. Chem. Insights* **13,** 1–5 (2018).
4. Reiss, P. D., Min, D. & Leung, M. Y. Working towards a consensus for antibody validation. *F1000Research* **3,** 1–3 (2014).
5. Acharya, P., Quinlan, A. & Neumeister, V. The ABCs of finding a good antibody: How to find a good antibody, validate it, and publish meaningful data. *F1000Research* **6,** 1–14 (2017).
6. Wardle, F. C. & Tan, H. A ChIP on the shoulder? Chromatin immunoprecipitation and validation strategies for ChIP antibodies. *F1000Research* **4,** 1–10 (2015).
7. Greenfield, E. A. *Antibodies: A laboratory manual. Second edition*. (Cold Spring Harbor Laboratory Press, 2014).
8. Kidder, B. L., Hu, G. & Zhao, K. ChIP-Seq: technical considerations for obtaining high-quality data. *Nat. Immunol.* **12,** 918–922 (2011).
9. Friguet, B., Chaffotte, A. F., Djavadi-Ohaniance, L. & Goldberg, M. E. Measurements of the true affinity constant in solution of antigen-antibody complexes by enzyme-linked immunosorbent assay. *J. Immunol. Methods* **77,** 305–319 (1985).
10. Myszka, D. G. Kinetic, equilibrium, and thermodynamic analysis of macromolecular interactions with BIACORE in *Methods in Enzymology* **323,** 325–340 (2000).
11. Neri, D., Montigiani, S. & Kirkham, P. M. Biophysical methods for the determination of antibody-antigen affinities. *Trends Biotechnol.* **14,** 465–470 (1996).
12. Glass, T. R., Ohmura, N. & Saiki, H. Least detectable concentration and dynamic

range of three immunoassay systems using the same antibody. *Anal. Chem.* **79,** 1954–1960 (2007).

13. Bee, C. *et al.* Exploring the dynamic range of the kinetic exclusion assay in characterizing antigen-antibody interactions. *PLoS One* **7,** e36261 (2012).
14. Drake, A. W. *et al.* Biacore surface matrix effects on the binding kinetics and affinity of an antigen/antibody complex. *Anal. Biochem.* **429,** 58–69 (2012).
15. Heinrich, L., Tissot, N., Hartmann, D. J. & Cohen, R. Comparison of the results obtained by ELISA and surface plasmon resonance for the determination of antibody affinity. *J. Immunol. Methods* **352,** 13–22 (2010).
16. Reverberi, R. & Reverberi, L. Factors affecting the antigen-antibody reaction. *Blood Transfus.* **5,** 227–240 (2007).
17. Dimitriadis, G. J. Effect of detergents on antibody-antigen interaction. *Anal. Biochem.* **98,** 445–451 (1979).
18. Wang, Y. V. *et al.* Quantitative analyses reveal the importance of regulated Hdmx degradation for P53 activation. *Proc. Natl. Acad. Sci. USA* **104,** 12365–12370 (2007).
19. Janes, K. A. An analysis of critical factors for quantitative immunoblotting. *Sci. Signal.* **8,** 1–12 (2015).
20. Dixon, A. S. *et al.* NanoLuc complementation reporter optimized for accurate measurement of protein interactions in cells. *ACS Chem. Biol.* **11,** 400–408 (2016).
21. Schwinn, M. K. *et al.* CRISPR-mediated tagging of endogenous proteins with a luminescent peptide. *ACS Chem. Biol.* **13,** 467–474 (2018).
22. Oh-hashi, K., Furuta, E., Fujimura, K. & Hirata, Y. Application of a novel HiBiT peptide tag for monitoring ATF4 protein expression in Neuro2a cells. *Biochem. Biophys. Reports* **12,** 40–45 (2017).
23. Sasaki, M. *et al.* Development of a rapid and quantitative method for the analysis of

viral entry and release using a NanoLuc luciferase complementation assay. *Virus Res.* **243,** 69–74 (2018).

24. Brizzard, B. Epitope tagging. *Biotechniques* **44,** 693–695 (2008).

25. Maue, R. A. Understanding ion channel biology using epitope tags: Progress, pitfalls, and promise. *J. Cell. Physiol.* **213,** 618–625 (2007).

26. Kanca, O., Bellen, H. J. & Schnorrer, F. Gene tagging strategies to assess protein expression, localization, and function in drosophila. *Genetics* **207,** 389–412 (2017).

27. Partridge, E. C., Watkins, T. A. & Mendenhall, E. M. Every transcription factor deserves its map: Scaling up epitope tagging of proteins to bypass antibody problems. *BioEssays* **38,** 801–811 (2016).

28. Savic, D. *et al.* CETCh-seq: CRISPR epitope tagging ChIP-seq of DNA-binding proteins. *Genome Res.* **25,** 1581–1589 (2015).

29. Yang, H. *et al.* One-step generation of mice carrying reporter and conditional alleles by CRISPR/Cas-mediated genome engineering. *Cell* **154,** 1370–1379 (2013).

30. Dewari, P. S. *et al.* An efficient and scalable pipeline for epitope tagging in mammalian stem cells using Cas9 ribonucleoprotein. *eLife* **7,** e35069 (2018).

31. Einhauer, A. & Jungbauer, A. The FLAG™ peptide, a versatile fusion tag for the purification of recombinant proteins. *J. Biochem. Biophys. Methods* **49,** 455–465 (2001).

32. Hopp, T. P. *et al.* A short polypeptide marker sequence useful for recombinant protein identification and purification. *Bio/Technology* **6,** 1204–1210 (1988).

33. Field, J. *et al.* Purification of a RAS-responsive adenylyl cyclase complex from *Saccharomyces cerevisiae* by use of an epitope addition method. *Mol. Cell. Biol.* **8,** 2159–2165 (1988).

34. Southern, J. a, Young, D. F., Heaney, F., Baumgartner, W. K. & Randall, R. E.

Identification of an epitope on the P and V proteins of simian virus 5 that distinguishes between two isolates with different biological characteristics. *J. Gen. Virol.* **72,** 1551–1557 (1991).

35. Fujii, Y. *et al.* PA tag: A versatile protein tagging system using a super high affinity antibody against a dodecapeptide derived from human podoplanin. *Protein Expr. Purif.* **95,** 240–247 (2014).

36. Bastin, P., Bagherzadeh, A., Matthews, K. R. & Gull, K. A novel epitope tag system to study protein targeting and organelle biogenesis in *Trypanosoma brucei*. *Mol. Biochem. Parasitol.* **77,** 235–239 (1996).

37. Lobbestael, E. *et al.* Immunohistochemical detection of transgene expression in the brain using small epitope tags. *BMC Biotechnol.* **10,** 1–10 (2010).

38. Brinkman, A. B. & Stunnenberg, H. G. Strategies for epigenome analysis in *Epigenomics* 3–18 (Springer Netherlands, 2009).

39. Luo, S., Wehr, N. B. & Levine, R. L. Quantitation of protein on gels and blots by infrared fluorescence of Coomassie blue and Fast Green. *Anal. Biochem.* **350,** 233–238 (2006).

40. Collett, M. S. & Erikson, R. L. Protein kinase activity associated with the avian sarcoma virus *src* gene product. *Proc. Natl. Acad. Sci. USA* **75,** 2021–2024 (1978).

41. Zhang, L., Rayner, S., Katoku-Kikyo, N., Romanova, L. & Kikyo, N. Successful co-immunoprecipitation of Oct4 and Nanog using cross-linking. *Biochem. Biophys. Res. Commun.* **361,** 611–614 (2007).

42. Pollard, T. D. A guide to simple and informative binding assays. *Mol. Biol. Cell* **21,** 4061–4067 (2010).

43. Cristea, I. M., Williams, R., Chait, B. T. & Rout, M. P. Fluorescent proteins as proteomic probes. *Mol. Cell. Proteomics* **4,** 1933–1941 (2005).

44. Burbelo, P. D., Goldman, R. & Mattson, T. L. A simplified immunoprecipitation method for quantitatively measuring antibody responses in clinical sera samples by using mammalian-produced Renilla luciferase-antigen fusion proteins. *BMC Biotechnol.* **5,** 1–10 (2005).

45. Zhang, H., Williams, P. S., Zborowski, M. & Chalmers, J. J. Binding affinities/avidities of antibody–antigen interactions: Quantification and scale-up implications. *Biotechnol. Bioeng.* **95,** 812–829 (2006).

46. Goldberg, M. E. & Djavadi-Ohaniance, L. Methods for measurement of antibody/antigen affinity based on ELISA and RIA. *Curr. Opin. Immunol.* **5,** 278–281 (1993).

47. Kimura, H., Hayashi-Takanaka, Y., Goto, Y., Takizawa, N. & Nozaki, N. The organization of histone H3 modifications as revealed by a panel of specific monoclonal antibodies. *Cell Struct. Funct.* **33,** 61–73 (2008).

48. Braunstein, M., Rose, A. B., Holmes, S. G., Allis, C. D. & Broach, J. R. Transcriptional silencing in yeast is associated with reduced nucleosome acetylation. *Genes Dev.* **7,** 592–604 (1993).

49. Gentsch, G. E., Patrushev, I. & Smith, J. C. Genome-wide snapshot of chromatin regulators and states in *Xenopus* embryos by ChIP-Seq. *J. Vis. Exp.* **96,** 1–10 (2015).

50. Dahl, J. A. & Collas, P. Q^2 ChIP, a quick and quantitative chromatin Immunoprecipitation assay, unravels epigenetic dynamics of developmentally regulated genes in human carcinoma cells. *Stem Cells* **25,** 1037–1046 (2007).

51. Wegner, G. J., Lee, H. J. & Corn, R. M. Characterization and optimization of peptide arrays for the study of epitope−antibody interactions using surface plasmon resonance imaging. *Anal. Chem.* **74,** 5161–5168 (2002).

52. Firsov, D. *et al.* Cell surface expression of the epithelial Na channel and a mutant

causing Liddle syndrome: A quantitative approach. *Proc. Natl. Acad. Sci. USA* **93,** 15370–15375 (1996).

53. Park, S. H. *et al.* Generation and application of new rat monoclonal antibodies against synthetic FLAG and OLLAS tags for improved immunodetection. *J. Immunol. Methods* **331,** 27–38 (2008).

54. Honey, S. A novel multiple affinity purification tag and its use in identification of proteins associated with a cyclin-CDK complex. *Nucleic Acids Res.* **29,** 1–9 (2001).

55. Domanski, M. *et al.* Improved methodology for the affinity isolation of human protein complexes expressed at near endogenous levels. *Biotechniques* **0,** 1713–1723 (2012).

56. Hernan, R., Heuermann, K. & Brizzard, B. Multiple epitope tagging of expressed proteins for enhanced detection. *Biotechniques* **28,** 789–93 (2000).

57. Kemmer, G. & Keller, S. Nonlinear least-squares data fitting in Excel spreadsheets. *Nat. Protoc.* **5,** 267–281 (2010).

58. Glass, T. R. *et al.* Development and characterization of new monoclonal antibodies specific for coplanar polychlorinated biphenyls. *Anal. Chim. Acta* **517,** 161–168 (2004).

59. Nishikori, S. *et al.* Broad ranges of affinity and specificity of anti-histone antibodies revealed by a quantitative peptide immunoprecipitation assay. *Journal of Molecular Biology* **424,** 391–399 (2012).

60. Hattori, T. *et al.* Recombinant antibodies to histone post-translational modifications. *Nat. Methods* **10,** 992–995 (2013).

61. Drake, A. W., Myszka, D. G. & Klakamp, S. L. Characterizing high-affinity antigen/antibody complexes by kinetic- and equilibrium-based methods. *Anal. Biochem.* **328,** 35–43 (2004).

62. Sosnick, T. R., Benjamin, D. C., Novotny, J., Seeger, P. A. & Trewhella, J. Distances

between the antigen-binding sites of three murine antibody subclasses measured using neutron and x-ray scattering. *Biochemistry* **31,** 1779–1786 (1992).

63. Zhang, L., Hernan, R. & Brizzard, B. Multiple tandem epitope tagging for enhanced detection of protein expressed in mammalian cells. *Mol. Biotechnol.* **19,** 313–322 (2001).

64. Marcon, E. *et al.* Assessment of a method to characterize antibody selectivity and specificity for use in immunoprecipitation. *Nat. Methods* **12,** 725–731 (2015).

65. Hakhverdyan, Z. *et al.* Rapid, optimized interactomic screening. *Nat. Methods* **12,** 553–560 (2015).

66. Wilkinson, K. D. Quantitative analysis of protein–protein interactions in *Methods in Molecular Biology* **261:** *Protein-Protein Interactions***,** 15–32 (Humana Press, 2004).

67. Link, V., Shevchenko, A. & Heisenberg, C. P. Proteomics of early zebrafish embryos. *BMC Dev. Biol.* **6,** 1–9 (2006).

Chapter 3.

Establishment of transgenic zebrafish lines with epitope-tagged *sox3* gene using CRISPR/Cas9-mediated knock-in approach

3-1 Introduction

The ability of precise and sequence-specific genome editing using the bacterial clustered regularly interspaced short palindromic repeats (CRISPR)–CRISPR-associated protein 9 (Cas9) system has revolutionized the genomic engineering and has equipped scientists with an unprecedented capability of modifying the genome of almost any kind of organisms[1,2]. The programable guide RNA (gRNA) complex which consists of CRISPR RNA (crRNA) and transactivating CRISPR RNA (tracrRNA) directs Cas9 endonuclease to the genomic target site to introduce double-strand breaks (DSBs) (Fig. 3-1)[3]. The 20-base spacer region of crRNA can be user-defined in order to specify the target genome region while the tracrRNA is a universal component. The tracrRNA hybridizes to the complementary region of the crRNA, and the combined crRNA and tracrRNA interacts with the Cas9 endonuclease forming the ribonucleoprotein (RNP) complex. Delivery of these CRISPR components in the form of RNP provides optimal genome editing efficiency and reduces unwanted off-target cutting[4].

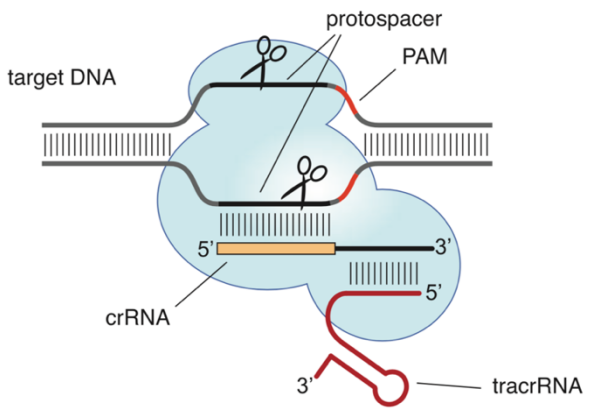


Figure 3-1. CRISPR/Cas9 system, ribonucleoprotein complex consists of Cas9 protein and tracrRNA:crRNA duplex.

In the natural CRISPR/Cas9 system, the target site consists of the protospacer that is complementary to the spacer region of the crRNA, and the protospacer adjacent motif (PAM). PAM is the sequence feature recognized by a specific CRISPR endonuclease once it is hybridized with a crRNA. For Cas9 endonuclease, PAM sequence is NGG where N can be any nucleotide[3,4]. As the DNA is a double helix consists of sense and anti-sense strands, RNP complex can be landed on to either the sense strand or the anti-sense strand depending on the realisation of the above mentioned necessities. Thus, the strand that is complementary to the crRNA spacer element is termed the 'target' strand whilst the strand that contains the PAM site which is complementary to the target strand is termed the 'non-target' strand.

Upon recognition of the target DNA region, Cas9 mediates the cleavage of target DNA upstream of PAM to create a DSB within the protospacer. This DSB is then healed by cellular repair machinery by either non-homologous end joining (NHEJ) pathway or the homology-directed repair (HDR) pathway[5,6]. In higher eukaryotic cells, the NHEJ pathway is predominant; however, it is often imprecise because it simply re-joins the two broken ends together which causes disruptive insertions and deletions (indels) at the target loci. If these indels occur in a coding exon that allows the efficient creation of gene knockouts[7]. On the other hand, for HDR pathway the presence of a sister chromatid which is naturally provided during the G2 and S phases of the cell cycle, or an exogenous donor DNA template is required. Although HDR repair is relatively less efficient, we can insert a novel genetic material (knock-in) precisely at the cleavage site by supplying an exogenous repair template that contains homology arms flanking the cleavage site. These template sequences can be designed for precise single nucleotide polymorphism (SNP) exchange[8], the knock-in of epitope tags such as V5[9,10], HA[11] and loxP sites[10,12], as well as for larger fluorescent protein tags[10]. However, it is of note that the HDR-mediated knock-in (KI) mechanism can also be error-prone yielding targeted insertions along with the unwanted indels[13].

The use of single-strand DNA (ssDNA) as the donor DNA template has recently been shown to be more effective in HDR-based genome editing due to its less cytotoxicity and higher insertion efficiency to the genome compared to the double-strand DNA (dsDNA) templates such as plasmids[10,14,15]. While single-strand oligodeoxynucleotides (ssODNs) of less than 200 bases can be easily synthesised and available commercially, the generation of long ssDNAs (lssDNAs) of more than 1 kb length is cumbersome but feasible using strategies such as *Easi*-CRISPR or using nicking enzymes[14,15]. Based on gRNA placement, ssDNAs can be termed as 'target' or 'non-target' strands. Non-target strand corresponds to the strand that is not bound by the gRNA, which contains the PAM sequence and conversely, the target strand corresponds to the strand that is bound by the gRNAs.

For efficient HDR insertion, the rational design of the ssDNA donor template is vital and one has to determine empirically several factors such as the orientation of the ssDNA, homology arm length and symmetry or asymmetry of homology arm lengths that would best perform for each case[16]. Apparently, the HDR efficiency is difficult to predict as it varies with the target site or the locus[13].

Another key consideration for efficient HDR insertion is the distance between the mutation or insertion (modification) and the Cas9 cut site (3 bp upstream from PAM). The most efficient position for a modification should be located <15 nucleotides (nt) and ideally <10 nt away from the cut site, as at a distance of 20 nt away from the cut site, the efficiency drops to 20-30% of the maximum as observed in murine cells[16,17].

HDR-based exogenous sequence integrations using CRISPR/Cas9 technology in zebrafish were found to be feasible, yet considerable improvements are required to obtain precise and heritable genetic alterations[18,19]. Using TALEN-based methods, successful integration and relatively high rate of germline transmission of small modifications as well as longer sequences such as fluorescent proteins in zebrafish have been reported[20–22]. However,

the use of TALENs has not been widely incorporated in knock-in experiments due to the greater difficulty of using TALENs than producing gRNAs. Several studies have tested ways and means to improve the efficiency of HDR-based CRISPR genome editing, and one such approach is to suppress the NHEJ pathway for instance by treating with chemicals such as Scr7 that can inhibit the components of the NHEJ pathway[23], or to enhance the HDR pathway by treating with chemicals such as RS-1[24]. Although these approaches were found to be effective for knock-ins in cultured cells and in mice[24–26], a low or no effect has been observed in zebrafish studies[13,27].

This study aimed to generate transgenic zebrafish lines by knock-in a composite of epitope tags (~200 bp in length) to the *sox3* locus. The careful selection of highly efficient crRNAs, and the rational design of the ssDNA donor template structure were the key considerations in this study.

3-2 Results and Discussion

3-2-1 Study design.

In this study, it was attempted to precisely knock-in a composite of epitope tags (~200 nt length), containing an epitope tag FLAG or PA in its trimeric form, followed by a TEV (tobacco etch virus) protease cleavage site, a biotin acceptor domain (Bio tag) and, most C-terminally, the HiBiT peptide tag, to the 3' end of the coding sequence of the *sox3* gene using the CRISPR/Cas9 genome editing tool. In order to knock-in the exogenous epitope tag sequence to the *sox3* gene using homology-dependent repair (HDR) mechanism after the CRISPR-mediated DSB, a ssDNA donor template was used.

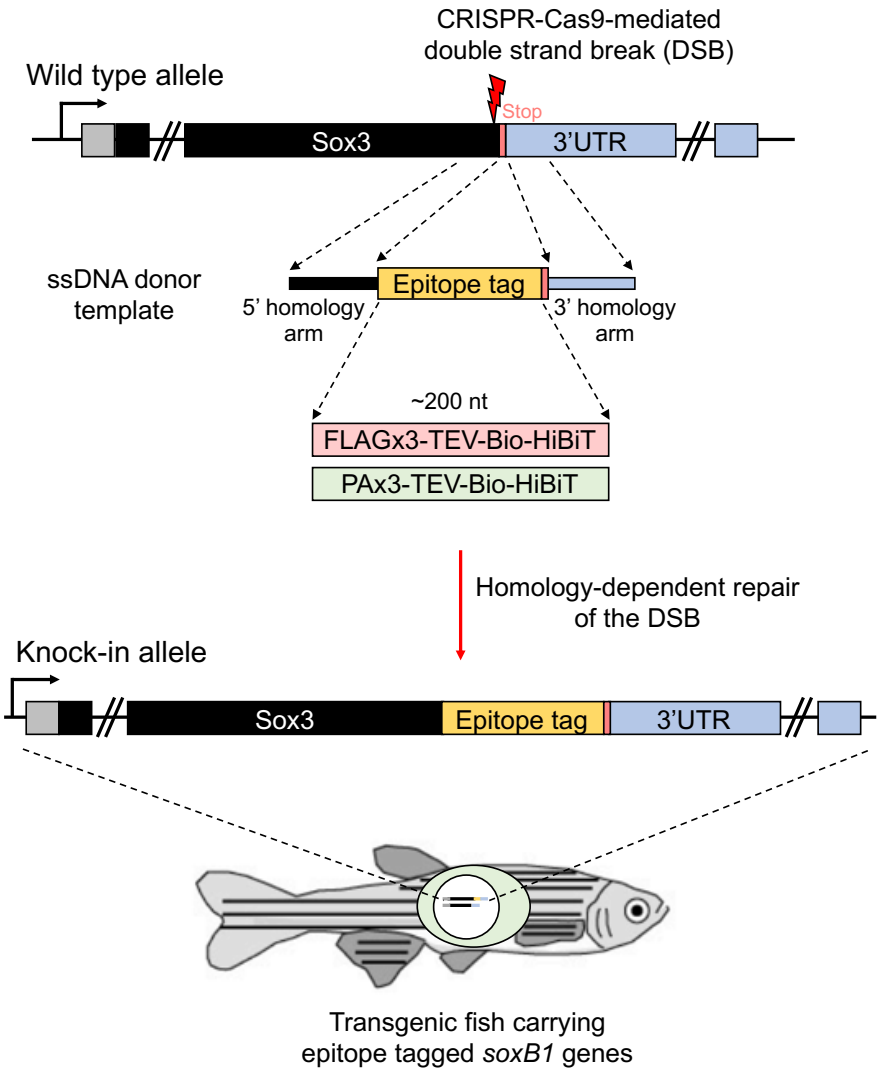


Figure 3-1. Study design.

3-2-2 CRISPR-genome-editing workflow.

CRISPR system used in this study includes two synthetic RNA oligonucleotides, a target-specific crRNA, and a universal tracrRNA, that are chemically-modified and length optimized variants of the native guide RNAs[4]. These two molecules form the guide RNA (gRNA) complex in a 1:1 ratio. The Cas9 protein used here is a high fidelity recombinant of *Streptococcus pyogenes* Cas9 that shows reduced off-target effects[28]. The gRNA complex and the Cas9 protein form the RNP complex in a 1:1 ratio. For genome editing, a mixture of the RNP complex and the ssDNA donor template is microinjected to the cytoplasm/ooplasm of the zebrafish embryos.

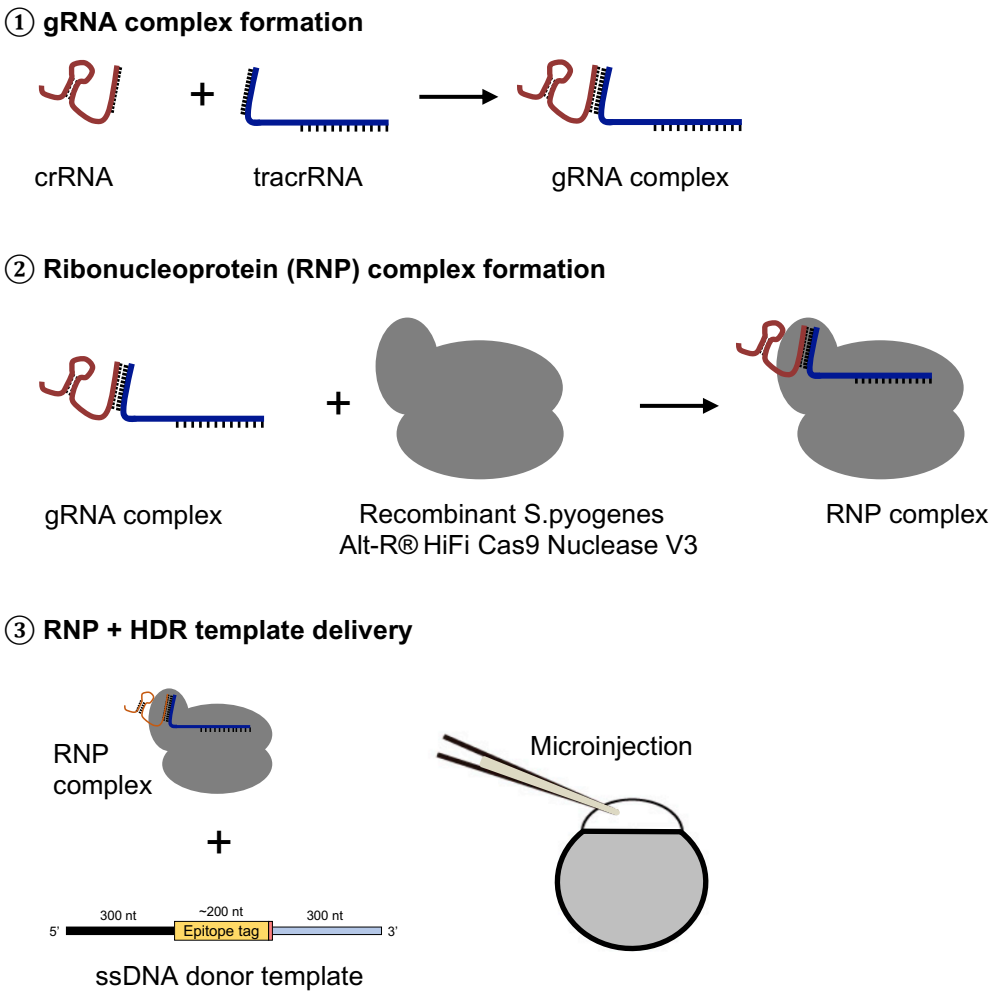


Figure 3-2. CRISPR-genome-editing workflow. The steps of crRNA:tracrRNA annealing, RNP complex formation with recombinant Cas9 protein, and embryo delivery are schematically outlined (adapted from Jacobi et al.[4]).

3-2-3 In-silico selection of efficient crRNA.

Candidate crRNA(s) were selected that would direct Cas9 to mediate DSB as close as possible to the stop codon of the *sox3* gene. Further, to select crRNAs that are predicted to have a high on-target cleavage efficiency with low off-target cleavage potential, several scoring algorithms provided in sg.idtdna.com, crispor.tefor.net and crispr.mit.edu were tried and by judging the on- and off-target scores, two candidate crRNAs were selected (Fig. 3-2A), one with a least of 2 nt distance between the stop codon (epitope tag insertion site) and the Cas9 cut site (crRNA_#1), while the other with a 43 nt distance (crRNA_#2).

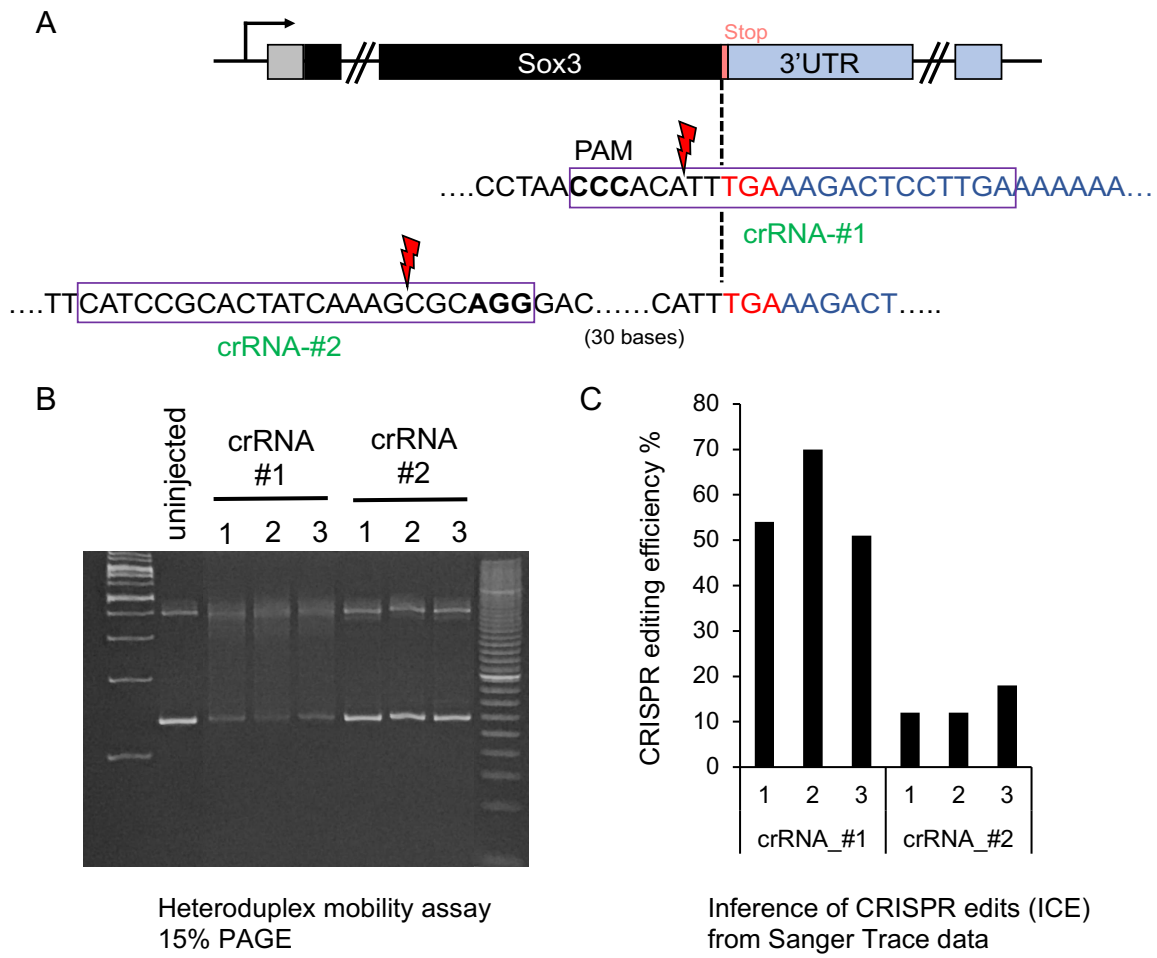


Figure 3-3. Determination of selected crRNA cleavage efficiency. (A) Candidate crRNA location and sequence (B) Heteroduplex mobility assay to evaluate the cleavage efficiency of crRNAs. A volume of 2 nL containing 3µM RNP complex was microinjected. Primers used for HMA are listed in Table 3-2. (C) Percentage of indel mutations by Inference of CRISPR Edits (ICE) analysis. Three replicates of embryo lysates per crRNA in both (B) and (C) analyses.

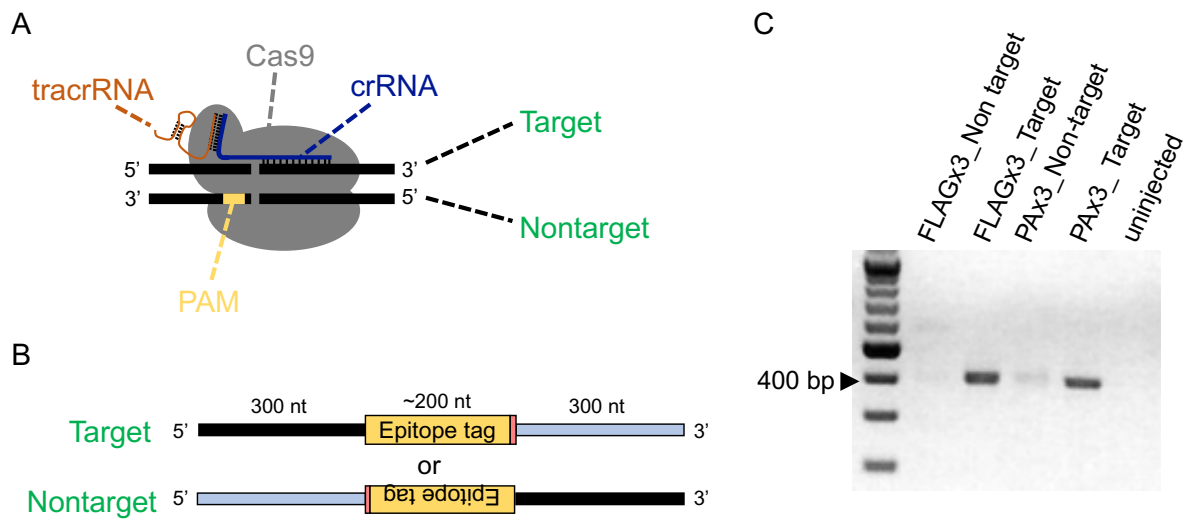
3-2-4 Evaluation of crRNA cleavage efficiency in vivo.

To examine the target site cleavage efficiency or the CRISPR activity of the selected two crRNAs, at first, heteroduplex mobility assay (HMA) was performed, in which a target region-specific PCR amplification is followed by polyacrylamide gel electrophoresis (PAGE). The NHEJ repair of CRISPR-induced DSBs creates indel mutations at the targeted genomic loci. Thus, when a target region-specific PCR amplification is performed, it results in a pool of PCR amplicons containing the intact wild-type sequence and the corresponding sequence with indel mutations. The subsequent denaturation followed by annealing of this pool of PCR-amplicons generates homoduplexes and heteroduplexes. These heteroduplexes can be separated from the homoduplexes by PAGE because the former migrates more slowly due to an opened single-strand configuration surrounding the mismatched region[29].

RNP complexes containing either crRNA_#1 or crRNA_#2 were microinjected into the cytoplasm of 1-cell stage embryos and genomic DNA was extracted after 24 hours post fertilization (hpf). When the cleavage efficiency was assessed by HMA assay, DNA smears which represents heteroduplexes derived from indels were observed for the CRISPR/Cas9 injected embryos in contrast to the uninjected control and the intensity of the smear was much more evident for crRNA_#1 than for crRNA_#2, indicating the efficient cleavage of the genome when using crRNA_#1 (Fig. 3-2B). Moreover, the percentage of genomes that have been modified with indels were identified using ICE (Inference of CRISPR Edits) software by submitting the sanger trace data derived from the PCR amplicons of the target region (Fig. 3-3C). Averages of approximately 60% and 14% of indel mutation rates were observed for crRNA_#1 and #2, respectively. Thus, crRNA_#1 was used during *sox3* gene editing procedures described below.

3-2-5 Selection of ssDNA orientation for efficient knock-in.

Because of the reported success of ssDNA donor templates during CRISPR-KI experiments compared to the dsDNA templates in mouse and rat experiments, ssDNA was chosen as the donor DNA templates[14,30]. As the lengths of DNA fragments encoding FLAGx3-TEV-Bio-HiBiT and PAx3-TEV-Bio-HiBiT are 204 bp and 243 bp, respectively, and total lengths of donor templates containing 5' and 3' homology arms are beyond the size limit of standard chemical DNA synthesis, lssDNAs were prepared from plasmids using nicking enzymes. In our initial design of donor templates, the composite of epitope tags was flanked by the 300-nt homology arms of *sox3* coding DNA sequence (CDS) and the *sox3* 3'UTR region in either side, expecting to insert the epitope tags at the 3' end of the *sox3* CDS (Fig. 3-1). To examine the effect of lssDNA strand orientation on KI efficiency, either the target (strand complementary to and bound by the crRNA) or the non-target lssDNA strand was microinjected into the cytoplasm of the one-cell stage embryos along with the CRISPR/Cas9 RNP complex. Next, genomic DNA was isolated from a group of 20 embryos of 1 dpf and KI allele specific PCR at the 3' junction of the insertion was performed. The target strand was found to be more efficient in KI for both FLAGx3 and PAx3 insertions (Fig. 3-4).



(For figure legend, see next page)

Figure 3-4. Selection of ssDNA orientation for efficient KI. (A) Definition of the target and non-target strands (B) Two ssDNA strand orientations examined (C) Agarose gel image showing the PCR amplicons of KI allele-specific PCR

3-2-6 High efficient knock-in events revealed by PCR.

The target strand lssDNA orientation, which yielded a higher KI efficiency, was used as the donor DNA template and microinjected into the cytoplasm of the one-cell stage zebrafish embryos along with the CRISPR/Cas9 RNP complex. A total of 20 embryos were examined individually for successful KI by KI-allele specific PCR at the 3' junction of the insertion. Interestingly, 19 embryos out of 20 injected F0 embryos (95%) were positive by PCR screening for successful KI (Fig. 3-5).

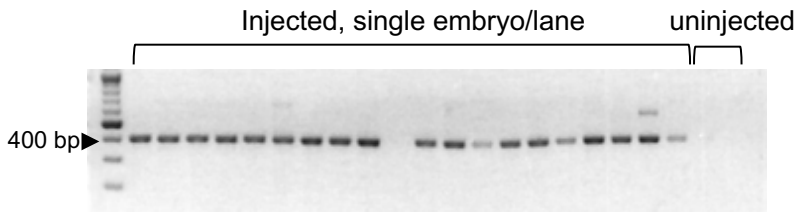


Figure 3-5. KI-allele specific PCR analysis of individual KI injected embryos.

Moreover, the survival of zebrafish embryos was examined after the injection of the CRISPR/Cas9 components along with the lssDNA donor template. A high rate of survival was observed even after 48 h post fertilization.

3-2-6 Effect of the asymmetric structure of the ssDNA donor template.

The structure of the ssDNA donor template plays an important role in determining the knock-in efficiency. Richardson *et al.* showed that the asymmetric target strand with 91 nt 5' homology arm and 36 nt 3' homology arm resulted in the highest HDR frequency[31]. They explain this observation by a model proposing that the Cas9 asymmetrically releases the 3' end of the cleaved DNA strand that is not complementary to the gRNA (non-target strand), and the ssDNA that can bind with this released strand which is the target ssDNA performs the

best. The shorter 3' homology arm facilitates the binding of the donor template to the 3' end of the released genomic DNA, and the homology-directed repair. Another model that supports the use of asymmetric ssDNA donor template with shorter 3' homology arm length is the protruding single-stranded 3' regions result from resection of double-strand breaks[32]. Another study that explored this model revealed supporting data for the use of asymmetric ssDNA donor template (97 nt in length) with shorter 3' homology arm length (30 nt) for efficient HDR irrespective of the strand orientation[16]. This study suggests that the model proposed by Richardson *et al.*[31] is not always true and one needs to examine both strand orientations for each particular study. Moreover, Yoshimi *et al.*[14] suggests the use of extended 5' homology arm to avoid the effect of possible exonuclease activity. Considering these points, two different lssDNA structures of the target strand were examined. The 5' homology arm of both the structures was 300-nt in length while the 3' homology arm length of each structure had 50-nt or 300-nt.

3-2-6 Screening of F0 fish to identify potential founders.

Target-strand of the ssDNA donor templates with either one of the two structures mentioned above was injected into the cytoplasm of the one-cell stage zebrafish embryos and raised to the adult-hood. Next, the F0 fish were in-crossed and pooled F1 embryos were analysed for the epitope tag integration events in the germline, in which both 5' and 3'-junctions of the insertion were examined by KI-allele specific PCR using tag-specific/genomic primer pairs. For the PCR positive samples, the PCR products were sequenced to characterize the integration events (Fig. 3-6).

Thirty-seven F0 fish were screened for the FLAGx3 tag insertion using lssDNA donor template with 300-nt 3' homology arm length and seven in-crossed pairs were found to be positive for KI-allele specific PCRs, indicating a minimum of 19% germline transmission

rate. Out of these seven pairs, one fish pair was identified with correct integration at both the 3' and 5'-junctions. By out-crossing this pair with wild-type fish, a male fish was isolated as the correct KI founder. The rest of the six fish pairs showed incorrect integration at the 5', 3' or both junctions where sequence duplications and indels were observed. For the FLAGx3 tag insertion using ssDNA donor template with 50-nt 3' homology arm length, 28 F0 fish were screened and eight in-crossed pairs were positive for KI-allele specific PCRs with a >29% germline transmission rate. Interestingly, 6 pairs showed correct integration at both the 5' and 3' junctions and by out-crossing with wild-type fish, individual founders could be isolated. Out-cross of the potential founder FLAGx3-50_#23 suggested germline mosaicism as shown by a mix of trace data for the 5'-junction amplifications.

Thirty fish were screened for the PAx3 tag insertion using ssDNA donor template with 300-nt 3' homology arm length and five in-crossed pairs were positive for KI-allele specific PCRs with a >17% germline transmission rate. One pair showed correct integration at both the 5' and 3' junctions; however, when out-crossed to the wild-type fish none were positive for the tag integration suggesting the very low percentage of KI allele positive germ cells. For the PAx3 tag insertion using ssDNA donor template with 50-nt 3' homology arm length, 47 F0 fish were screened and ten in-crossed pairs were positive for KI-allele specific PCRs with a >21% germline transmission rate. Three pairs showed correct integration at both the 5' and 3' junctions and by out-crossing with wild-type fish, individual founders were isolated. For all the four scenarios analysed, the majority of the incorrect insertions were found to be occurred at the 5'-junction, apparently large deletions, as there were no PCR amplicons obtained at this junction (Fig. 3-6, Table 3-1).

Several founders were further analysed by out-crossing and preforming KI-allele specific PCR at both 5' and 3' junctions as individual F1 embryos to examine the germline mosaicism. Interestingly, FLAGx3-50_#16 founder was identified as a biallelic founder as

observed by the individual F1 embryo analysis. For both 5' and 3' junction PCR amplifications, 19 embryos out of 20 were positive for the KI allele, suggesting that the majority of the germ cells of this founder seems homozygous for the epitope tag insertion. However, two embryos showed different lengths for the PCR product suggesting the occurrence of multiple KI events, one as a minor event (Fig. 3-7A, red arrows). Possibly, the genotype of the germ cells with this minor event can be one allele with the wild type, while the other with incorrect integration (Fig. 3-7A).

The PAx3-50_#21 founder also demonstrated biallelic behaviour for the knock-in as evidenced by all positive embryos for the PCR amplification at the 5'-junction. However, only 11 out of 20 embryos were positive for the PCR amplification of the 3'-junction suggesting one KI allele is associated with a large deletion (Fig. 3-7B).

Tag-3' homology arm length	Number of fish screened	Number of germline transmitted fish (F0 founders)	Rate of germline transmission	F0 founders with correct integration		Number of F1 embryos with correct integration
FLAGx3-300	37	7	19% (7/37)	1	#9	16% (8/50)
FLAGx3-50	28	8	29% (8/28)	6	#9	N.D.
					#16	**85% (17/20)**
					#19	5% (1/20)
					#20	N.D.
					#21	N.D.
					#22	N.D.
PAx3-300	30	5	17% (5/30)	1	#7	N.D.
PAx3-50	47	10	21% (10/47)	3	#21	**55% (11/20)**
					#25	N.D.
					#34	N.D.

(N.D.: Not determined)

Table 3-1. Summary of screened F0 founders.

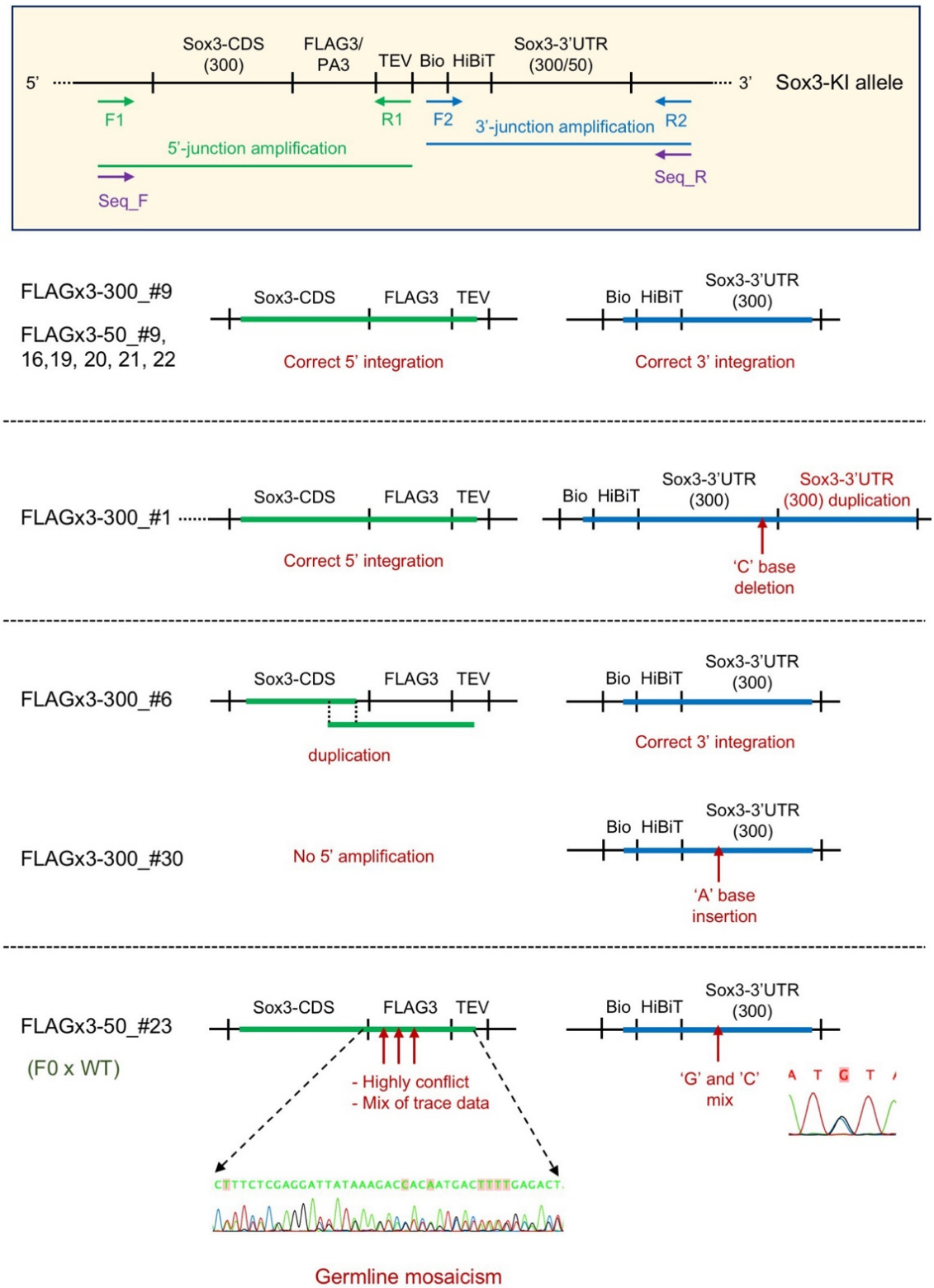


Figure 3-6. Screening of F0 fish to identify potential founders. Schematic representation of PCR amplification of the 5' and 3' junctions of epitope tag integrated into the *sox3* gene followed by PCR product sequencing. Position of PCR and sequencing primers are shown (upper box) and the sequence of each primer is listed in Table 3-2. The nomenclature of each possible founder is shown in the left panel. Sequence analysis of 5' (middle panel) and 3' (right panel) junction PCR products.

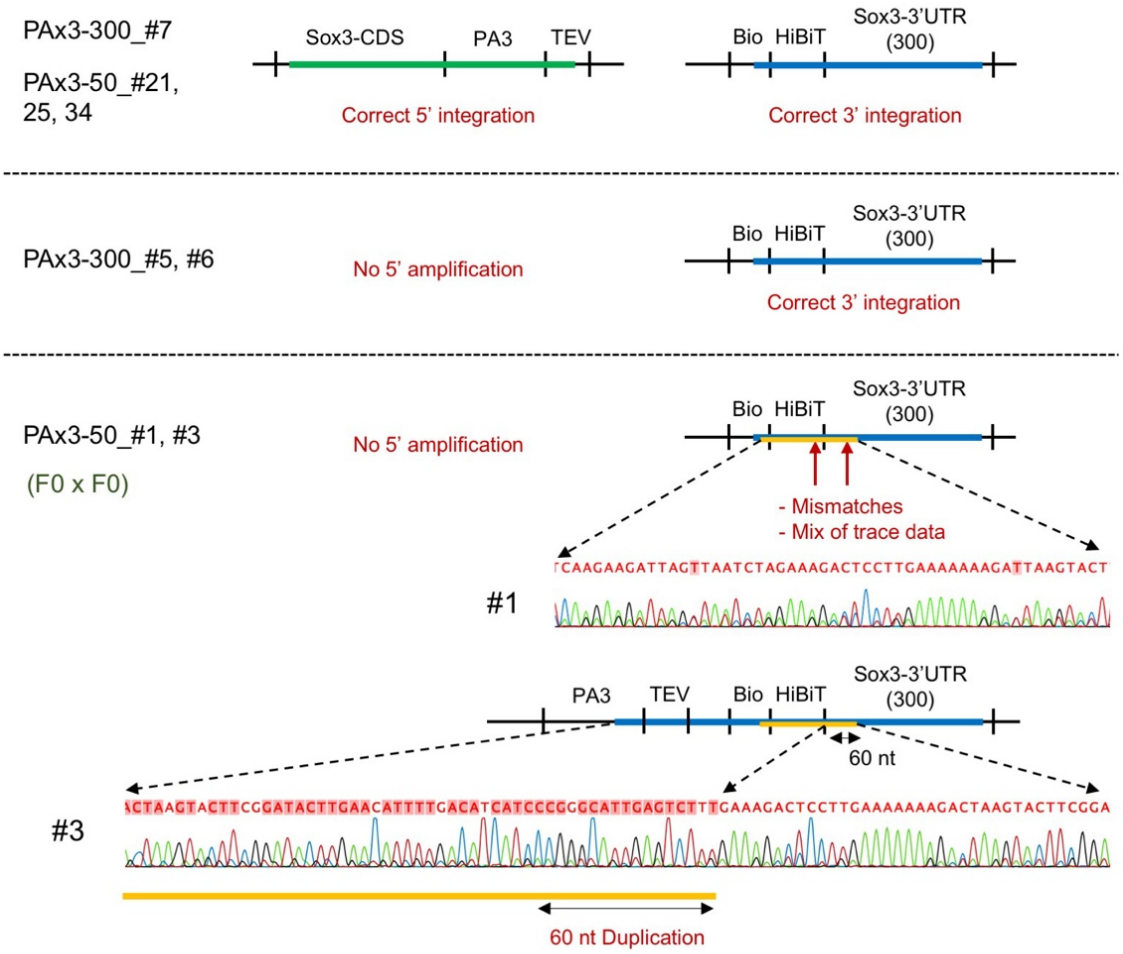


Figure 3-6. Screening of F0 fish to identify potential founders. The nomenclature of each possible founder is shown in the left panel. Sequence analysis of 5' (middle panel) and 3' (right panel) junction PCR products.

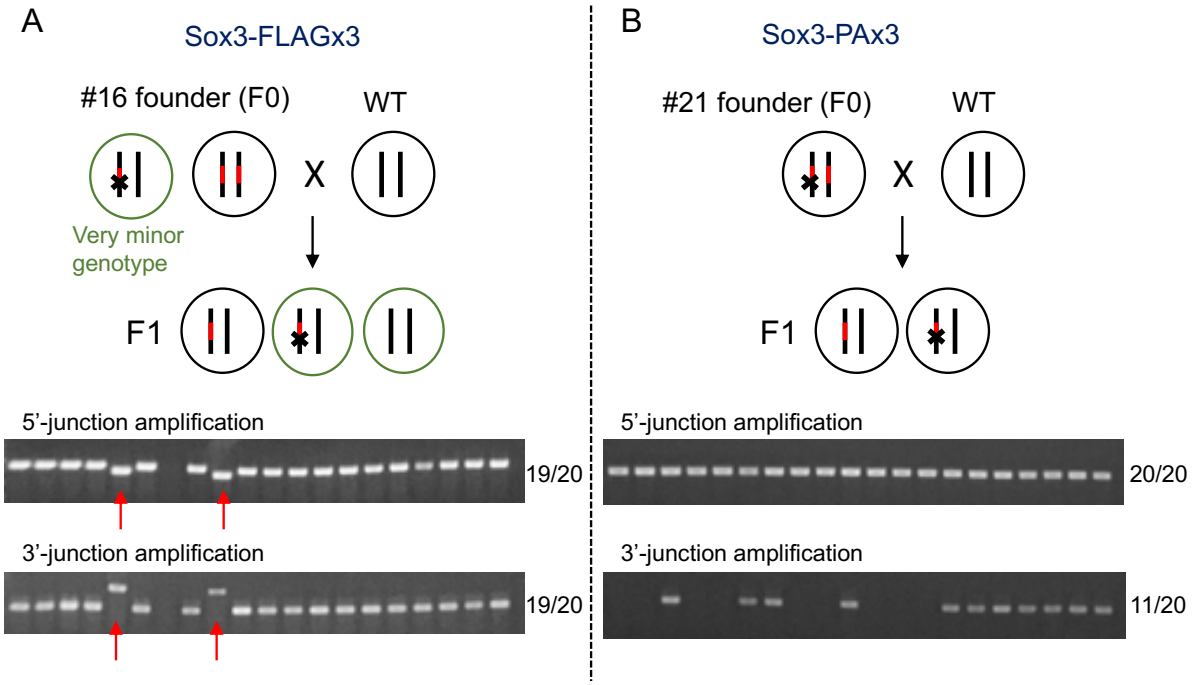


Figure 3-7. Possible genetic background of biallelic founders. Out-crosses of FLAGx3-50_#16 (A) and PAx3-50_#21 (B) founders and PCR amplification of 5' and 3' junctions of epitope tag integrated *sox3* gene of individual F1 progeny embryos. A total of 20 embryos per each founder were analysed. The number of PCR positive embryos per number of total embryos for each PCR amplification has shown in the left side of each agarose gel image. Possible genotypes of germ cells for each founder are depicted. Occurrence of multiple knock-in events marked in red arrows (A).

To find the abundance of the KI allele in founders, PCR was performed using primers that bind outside the homology arms. In this PCR, the amplification of the wild type allele is preferred due to the shorter length compared to the KI allele. Thus, some founders showed only the wild type allele amplification. For the founders with biallelic insertion, amplification of the KI allele was observed with varying intensities suggesting the variation of KI allele positive germ cell numbers in a given founder (Fig. 3-8, red arrowheads).

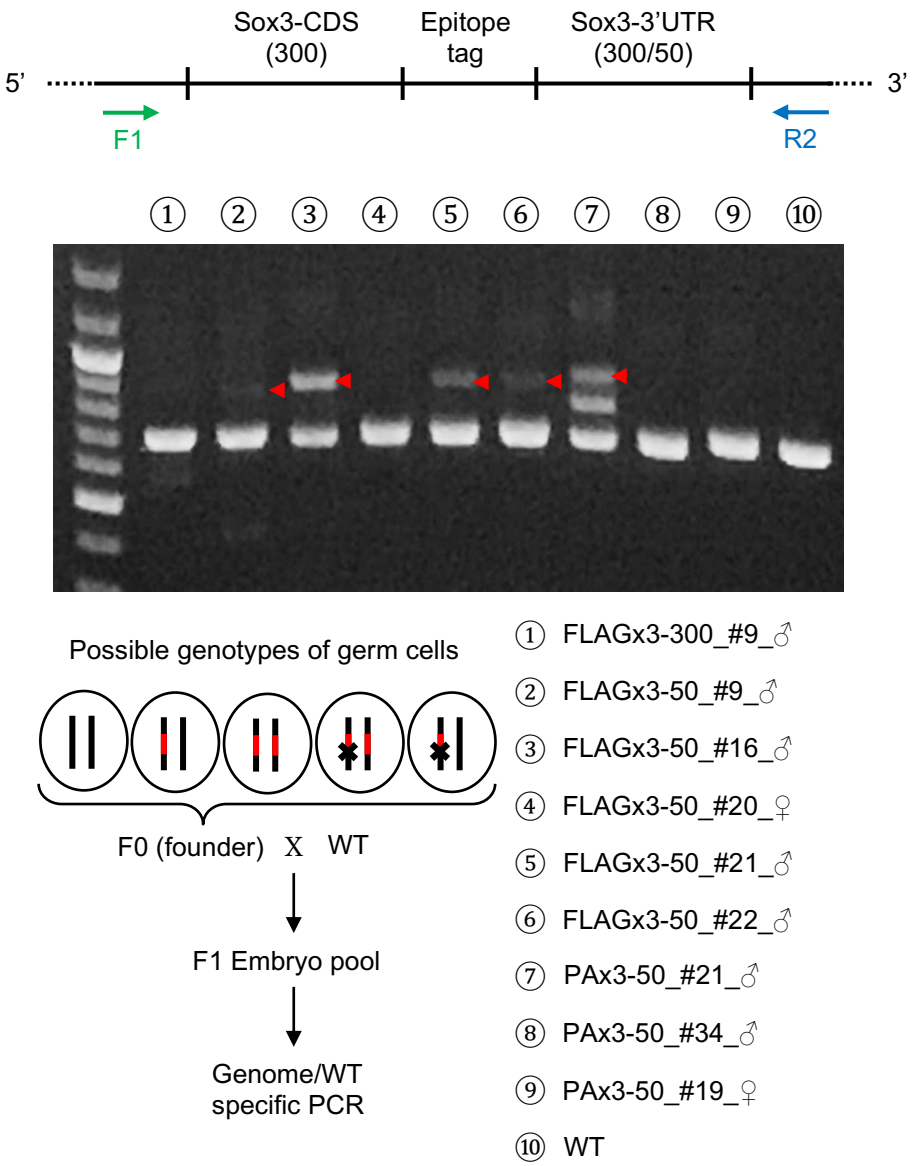


Figure 3-8. Genotypic mosaicism of the founders. PCR amplification of the knock-in allele using primers that bind outside of the homology arms. The sequence of each primer is listed in Table 3-2. The variation of KI allele positive germ cell numbers in a given founder marked in red arrowheads.

Precise genome editing using CRISPR/Cas9-mediated knock-in in zebrafish is yet challenging and requires improvements to increase efficiency and precision. Initial attempts to insert short sequences such as LoxP sites[12], restriction enzyme sites or single nucleotide substitutions[19] using single-stranded oligonucleotide (ssODN) donor templates reported the integration of the intended sequence, although along with the additional indels. Moreover, germline transmission of those integrations was not tested in most cases. However, Armstrong *et al.*[33] reported successful point mutation insertion and germline transmission rates of 2-4% in their study where they introduced point mutations corresponding to the identified mutations in amyotrophic lateral sclerosis (ALS) patients in the genes *tardbp* and *fus*.

Corresponding to this thesis study, although the insert length is relatively shorter, Hruscha *et al.*[11] demonstrated successful knock-in of one or two HA tags (27 nucleotides) using ssODN donor with homology arms between 30 and 50-nt. Although they could detect knock-in allele by PCR amplification of injected embryos as efficient as 45% and 70% for the *C13H9orf72* and *tardbp* loci, respectively, sequencing data revealed only 1.7% and 3.5% of correct integrations. As these are somatic modifications the chance of germline transmission rates would be much lower although they have not examined. Another attempt to insert V5 tag using ssODN templates with nearly 20-nt homology arms showed germline transmission of the tag integration for both *tcf21* and *tbx18* loci[34]. However, the results obtained for the tail clip DNA analysis of F1 fish for *tcf21* locus did not show precise integrations suggesting high germline mosaicism, while 21% of F1 fish for the *tbx18* locus showed precise tag integrations. Taken together, in both of the above-mentioned studies, the proportion of correctly-modified alleles was low. Therefore, the rate of precise and heritable integrations reported in this study is significantly high, although there is a necessity to provide further evidence of efficient integration for the other genes/loci.

3-2-6 Validation of epitope-tagged Sox3 expression.

Next, to confirm that the endogenous epitope-tagged Sox3 is expressed correctly without deleterious effects of KI and that the epitope tag is accessible for antibody binding, zebrafish whole-mount immunohistochemical analysis was performed using F1 embryos derived from the founders with biallelic insertion (Fig. 3-8). As expected, the FLAGx3 and PAx3 tagged Sox3 proteins were detected by anti-FLAG and anti-PA antibodies, respectively, similarly to anti-Sox3 antibody. Further, the comparable expression of the Sox3 proteins in the central nervous system of the KI fish with compared to the wild type confirmed that the epitope tags do not affect the expression or stability of the Sox3 protein.

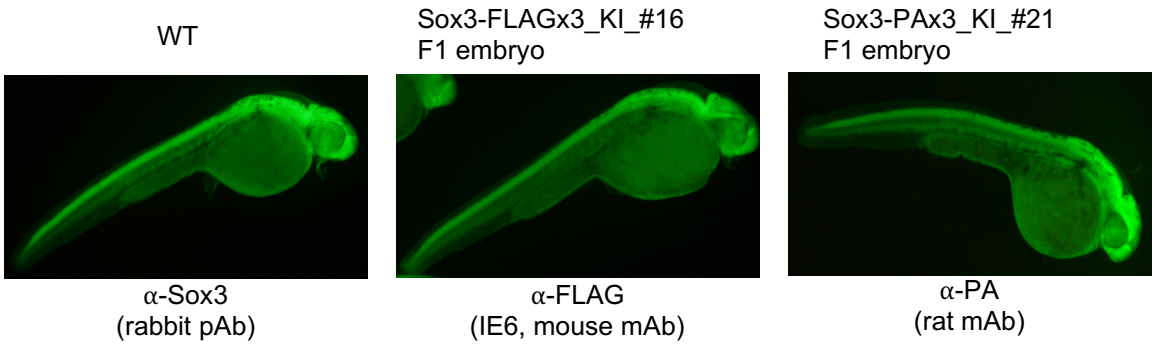


Figure 3-8. Whole-mount immunohistochemical analysis to examine endogenous epitope-tagged Sox3 protein expression. Each primary antibody used to stain zebrafish is depicted in the below of each image.

Further, the expression of endogenous epitope-tagged Sox3 proteins was validated using western blotting and HiBiT blotting (Fig. 3-9). Consistent with the whole-mount immunostaining data, anti-FLAG and anti-PA antibodies detected Sox3 proteins tagged with FLAGx3 and PAx3, respectively, with expected molecular sizes on the blot. Moreover, HiBiT peptide tag was also detected in both cases, suggesting their intact expression.

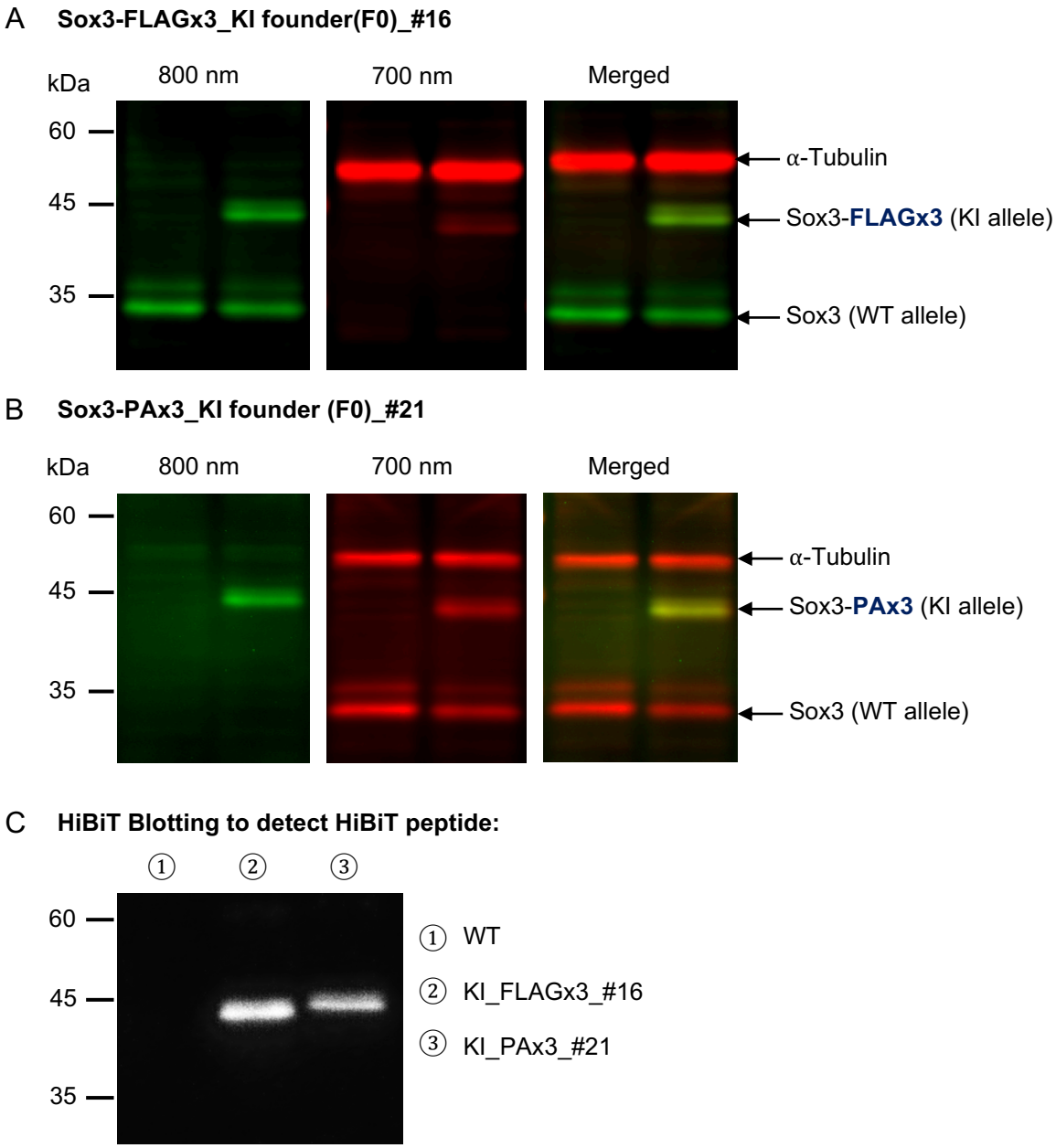


Figure 3-9. Western blotting and HiBiT blotting to examine endogenous epitope tagged Sox3 protein expression. (A, B) Two-colour western blot analysis of *sox3* knock-in embryos derived from the founders FLAGx3-50_#16 and PAx3-50_#21 with antibodies against Sox3, FLAG (A) or PA (B) and alpha-Tubulin. **(C)** HiBiT blotting using an excess amount of LgBiT protein and substrate to detect the C-terminal HiBiT peptide in *sox3* knock-in allele.

3-3 Materials and Methods

3-3-1 In-silico gRNA designing.

crRNAs were designed using the Integrated DNA Technologies, Inc. (IDT) CRISPR design tool (https://sg.idtdna.com/site/order/designtool/index/CRISPR_CUSTOM).

3-3-2 Zebrafish genomic DNA extraction.

At 1 dpf we categorized injected embryos as 'dead', 'malformed' or 'normally developing'. We performed genomic DNA extraction on a single embryo or a pool of normally developing embryos. Genomic DNA extraction buffer (200 mM NaCl, 10 mM Tris-HCl (pH 8), 10 mM EDTA, 1% Triton X-100, 200 µg/mL Proteinase K) was added to embryos and incubated at 55 ºC for several hours (2-3 hours) with occasional mixing until the embryos dissolved completely. Proteinase K was inactivated by heating the samples at 90 ºC for 10-12 min. When purified genomic DNA is required, RNaseA was added to the embryo lysate to have a final concentration of 0.1 mg/mL and incubated for 30 min. at room temperature. After vortexing to make the sample easy to load onto the column, DNA was purified using the NucleoSpin Gel and PCR Clean-up columns according to the manufacturer's protocol.

3-3-3 Heteroduplex Mobility Assay (HMA).

Primers were designed to amplify a region of approximately 150 bp around the stop codon of the *Sox3* gene including the gRNA target site (primer sequences are listed in Table 3-2). PCR conditions were as follows: 95 ºC for 30 s; 95 ºC for 15 s, 55 ºC for 30 s, 68 ºC for 10 s for 30 cycles; 68 ºC for 5 min followed with denaturation for 5 minutes at 95 ºC. PCR products were removed from the thermocycler and maintained at room temperature for at least 5 minutes allowing for annealing, before loading onto non-denaturing polyacrylamide

gel containing 15% acrylamide-bisacrylamide (29:1, w/w) in 1X Tris-borate-EDTA (TBE). After electrophoresis, the polyacrylamide gel was immersed in 0.5 µg/mL ethidium bromide solution for 40 minutes before visualization using Fusion imaging system (Vilber-Lourmat). For control purposes, a fraction of PCR products was resolved with ethidium bromide-containing 2% agarose gel.

3-3-4 Plasmid DNA construction for ssDNA donor template DNA preparation.

To prepare lssDNAs containing a composite of epitope tags-FLAGx3-TEV-Bio-HiBiT/ PAx3-TEV-Bio-HiBiT, flanked in either side with homology arms of Sox3 CDS and Sox3 3' UTR, at first, a 300 nt of Sox3 CDS upstream from the stop codon and 300 nt or 50 nt of Sox3 3'UTR region downstream from the stop codon were amplified from the purified zebrafish genomic DNA and cloned into pUC19 vector using EcoRI and PstI restriction sites. Taking this construct as a template, Sox3 CDS region was PCR amplified to have EcoRI and XhoI in either end (stop codon was removed) and Sox3 3'UTR to have XbaI and PstI in either ends. Additionally, the composite of epitope tag was constructed to have XhoI and XbaI in its either ends. Next, all three amplicons were ligated to generate pUC19_Sox3-CDS(300)_FLAGx3-TEV-Bio-HiBiT or PAx3-TEV-Bio-HiBiT_ Sox3-3'UTR(300/50).

3-3-5 lssDNA preparation.

lssDNAs were prepared using an LsODN Preparation Kit (Biodynamics Laboratory Inc., Tokyo, Japan) according to the manufacturer's protocol. A DNA fragment comprising a composite of epitope tags (either FLAGx3 or PAx3) and homology arms was cloned between the BspQI and the BbvCI sites of pLSODN-1. The resulting plasmid was digested with nicking endonucleases Nt.BspQI and Nb.BbvCI (in the case of FLAGx3) or HindIII and

Nt.BspQI (in the case of PAx3). Mixture of the nicked plasmid and denaturing gel-loading buffer was subjected to agarose gel electrophoresis. After electrophoresis, the gel was stained with 0.5 μg/mL ethidium bromide solution, which visualized three bands, a single-stranded DNA fragment comprising epitope tag composite and homology arms, a liner single-stranded vector DNA and a single-stranded circular whole-plasmid DNA. The band corresponding to a single-stranded DNA fragment was excised and extracted using NucleoSpin Gel and PCR Clean-up columns according to the manufacturer's protocol.

3-3-6 Zebrafish embryo microinjection.

Alt-R crRNA and tracrRNA (IDT) were resuspended in RNase-free 0.1xTE (1 mM Tris HCl (pH 7.0), 0.1 mM EDTA) to final concentrations of 100 μM each. To create a concentration of 3 μM gRNA solution, crRNA and tracrRNA were mixed 1:1 using Nuclease-free Duplex buffer (IDT) and heated at 95 ºC for 5 min. The complex was removed from heat and allowed to cool to room temperature. Next, Cas9 protein (Alt-R S.pyogenes HiFi Cas9 Nuclease V3, IDT) was diluted to a working concentration of 3 μM with RNase-free 0.1xTE or Cas9 working buffer (20 mM HEPES-KOH, 150 mM KCl, pH 7.5). The RNP complex (equal molar amounts of gRNA complex and Cas9 protein) was then incubated at 37 ºC for 10 min. and brought to the room temperature. The lssDNA template was mixed with RNP complex to a final concentration of 0.2 μM using 0.1xTE or Cas9 working buffer.

Following fertilization, 1.5 nL injection mix composed of 2 μM gRNA, 2 μM Cas9 protein, 0.2 μM ssDNA template and RNase-free 0.1xTE or Cas9 working buffer was microinjected into the cytoplasm of the one-cell-stage TL zebrafish embryos.

3-3-7 PCR assays to evaluate knock-in events and to screen F0 fish.

PCR was performed using primer pairs listed in Table 3-2. A 30 µL PCR reaction mix contained 0.75 U Taq DNA Polymerase (New England BioLabs), 1X Thermopolymerase buffer, 0.5 mM forward and reverse primer pairs, 0.2 mM dNTP mix, 1X sucrose red, and genomic DNA template (1-2 µL embryo lysate). The standard PCR condition was as follows: 95 ºC for 30 s; 95 ºC for 15 s, 62 ºC for 30 s, 68 ºC for 1 min/kb for 30 cycles; 68 ºC for 5 min.

Primer name	Sequence
Sox3_HMA_F1	ACTCCAGTCTACAGACCAGTC
Sox3_HMA_R1	TTCAAGTATCCGAAGTACTTAGTC
F1/Seq_F	GCGGGACTTCAGTACCCAATGA
R1	TGGAAGTACAGGTTCTCACGCG
F2	CGTGCCACTCGATCTTTTGAGC
R2/Seq_R	TACCCAATGATGTCCACGGCTC

Table 3-2. List of primers used for heteroduplex mobility assay, knock-in event evaluation, and F0 fish screening.

3-3-8 Whole-mount immunohistochemistry.

This method was adopted from Inoue *et al.*[35] F0 founders were out-crossed with wild type fish. After 24 hpf embryos were dechorionated and fixed in freshly prepared 4% formaldehyde for 1-2 hours at 4 ºC rotating. Embryos were washed with PBST three times, 5 min. per each wash at RT. Next, embryos were dehydrated by incubating with a successive dilutions of methanol in 1xPBS: 5 min in 25% (vol/vol) methanol; 5 min in 50% (vol/vol) methanol; 5 min in 75% (vol/vol) methanol and 5 min in 100% methanol[36]. Then, embryos were rehydrated in the opposite order with the same methanol series and washed with PBST three times, 5 min. per each wash at RT. Next, embryos were washed with 150 mM Tris-HCl (pH 9.0), 5 min. and incubated with 150 mM Tris-HCl (pH 9.0), 70 ºC, 15 min. for antigen retrieval followed by a PBST wash three times, 5 min. per each wash. On an ice bath,

embryos were rinsed with ice-cold dH_2O two times and permeabilized in prechilled acetone at -20 ºC, 20 min. Embryos were rinsed with ice-cold dH_2O two times to remove acetone. Next, embryos were washed with PBST three times, 5 min. per each wash at RT and blocked with a blocking buffer (10% Normal Goat Serum/PBT) at 4 ºC, 3 hrs, with rotating. Embryos were transferred into a 24-well plate and incubated in 1% Normal Goat Serum/PBT (1.0 M Sodium phosphate buffer [pH 7.4], 0.8% Triton X-100) with primary antibody at 4 ºC, 3 days with rocking agitation followed by washing with PBT five times, 1 hour each at RT. Primary antibodies were anti-Sox3 (rabbit pAb, GTX132494, GeneTex), anti-FLAG (mouse mAb, IE6, Wako), anti-PA (rat mAb, NZ-1, Wako) at final concentration of 0.5 ng/µL. Next, embryos were incubated in 1% Normal Goat Serum/PBT with secondary antibody (anti-rabbit IgG Alexa 488, anti-mouse IgG Alexa 488, anti-rat IgG Alexa 488) final concentration of 4 µg/mL at 4 ºC, 2 days with rocking agitation avoiding light followed by washing with PBT five times, 5 min. each at RT. Finally, embryos were mount with a gradient of glycerol in 1xPBS: 20 min in 25% (vol/vol) glycerol; 20 min in 50% (vol/vol) glycerol; 20 min in 75% (vol/vol) glycerol and viewed the staining under a fluorescent microscope.

3-3-9 Western blotting and HiBiT blotting to test endogenous epitope-tagged Sox3 expression.

F1 progeny embryos were collected by out-crossing the Sox3-FLAGx3-KI_#16 and PAx3-KI_#21 F0 founders with wild-type fish and raised until the 70-80% epiboly stage. Next, embryos were dechorionated enzymatically with pronase (2 mg/mL), deyolked as described by Link *et al*. and mixed with SDS sample buffer containing 50 mM DTT. Embryo lysates were then denatured by heating at 75 ºC for 5 min and proteins were separated by SDS-PAGE followed by transfer to nitrocellulose membranes (BioTrace NT, Pall

Corporation) using a TransBlot Cell (Bio-Rad).

To examine the expression of Sox3, alpha-Tubulin, and FLAG or PA tag using two-color detection, the membrane was treated as follows. First, the membrane was blocked with Odyssey blocking buffer (1:1 diluted with TBS), rinsed with TBST buffer (TBS and 0.05% Tween-20), and membrane strips were separated as FLAGx3-KI_#16 and PAx3-KI_#21 before incubating with primary antibodies. Both strips were probed with anti-Sox3 (GTX132494, GeneTex), anti-alpha Tubulin (B-5-1-2, Sigma) and each with anti-FLAG (M2, Sigma) or anti-PA (NZ-1, Wako) accordingly at 0.5 μg/mL final concentration except anti-alpha Tubulin in which final concentration was 1 μg/mL. Antibodies were diluted in Can Get Signal solution 1 (Toyobo). In the FLAGx3-KI_#16 membrane, Sox3 antibody was detected with goat anti-rabbit IgG-IRDye800, and alpha Tubulin and FLAG antibodies were detected with goat anti-mouse IgG-IRDye680 while in the PAx3-KI_#21 membrane, Sox3 antibody was detected with goat anti-rabbit IgG-CF680, alpha Tubulin antibody with goat anti-mouse IgG-IRDye680 and PA antibody with goat anti-rat IgG-IRDye800. All the secondary antibodies were diluted in Odyssey blocking buffer (1:1 diluted with TBS) containing 0.1% Tween-20 and 0.01% SDS. Images were acquired in both 700 and 800 nm fluorescent channels using an Odyssey CLx infrared imaging system (LI-COR Biosciences).

To detect the HiBiT peptide, the protein-transferred membranes were incubated in TBST for 30 min, and this medium was then replaced with Nano-Glo HiBiT blotting reagent containing LgBiT protein (Promega). After 1 hour of incubation at RT, the substrate furimazine was added, and the incubation was continued for another 5 min. The blot was imaged using a chemiluminescence imager with a CCD camera (Fusion, Vilber Lourmat).

3-4 References

1. Cong, L. *et al.* Multiplex genome engineering using CRISPR/Cas systems. *Science* **339,** 819–823 (2013).
2. Mali, P., Esvelt, K. M. & Church, G. M. Cas9 as a versatile tool for engineering biology. *Nat. Methods* **10,** 957–963 (2013).
3. Jinek, M. *et al.* A Programmable Dual-RNA – Guided DNA Endonuclease in Adaptive Bacterial Immunity. *Science* **337,** 816–822 (2012).
4. Jacobi, A. M. *et al.* Simplified CRISPR tools for efficient genome editing and streamlined protocols for their delivery into mammalian cells and mouse zygotes. *Methods* **121–122,** 16–28 (2017).
5. Kakarougkas, A. & Jeggo, P. A. DNA DSB repair pathway choice: An orchestrated handover mechanism. *Br. J. Radiol.* **87,** (2014).
6. Mao, Z., Bozzella, M., Seluanov, A. & Gorbunova, V. Comparison of nonhomologous end joining and homologous recombination in human cells. *DNA Repair* **7,** 1765–1771 (2008).
7. Shalem, O. *et al.* Genome-Scale CRISPR-Cas9 Knockout Screening in Human Cells. *Science* **343,** 84–87 (2014).
8. González, F. *et al.* An iCRISPR platform for rapid, multiplexable, and inducible genome editing in human pluripotent stem cells. *Cell Stem Cell* **15,** 215–226 (2014).
9. Burg, L. *et al.* Internal epitope tagging informed by relative lack of sequence conservation. *Sci. Rep.* **6,** 36986 (2016).
10. Yang, H. *et al.* One-Step Generation of Mice Carrying Reporter and Conditional Alleles by CRISPR/Cas-Mediated Genome Engineering. *Cell* **154,** 1370–1379 (2013).
11. Hruscha, A. *et al.* Efficient CRISPR/Cas9 genome editing with low off-target effects

in zebrafish. *Development* **140,** 4982–4987 (2013).

12. Chang, N. *et al.* Genome editing with RNA-guided Cas9 nuclease in Zebrafish embryos. *Cell Res.* **23,** 465–472 (2013).
13. Boel, A. *et al.* CRISPR/Cas9-mediated homology-directed repair by ssODNs in zebrafish induces complex mutational patterns resulting from genomic integration of repair-template fragments. *Dis. Model. Mech.* **11,** dmm035352 (2018).
14. Yoshimi, K. *et al.* ssODN-mediated knock-in with CRISPR-Cas for large genomic regions in zygotes. *Nat. Commun.* **7,** 10431 (2016).
15. Miura, H., Quadros, R. M., Gurumurthy, C. B. & Ohtsuka, M. Easi-CRISPR for creating knock-in and conditional knockout mouse models using long ssDNA donors. *Nat. Protoc.* **13,** 195–215 (2018).
16. Liang, X., Potter, J., Kumar, S., Ravinder, N. & Chesnut, J. D. Enhanced CRISPR/Cas9-mediated precise genome editing by improved design and delivery of gRNA, Cas9 nuclease, and donor DNA. *J. Biotechnol.* **241,** 136–146 (2017).
17. Paquet, D. *et al.* Efficient introduction of specific homozygous and heterozygous mutations using CRISPR/Cas9. *Nature* **533,** 125–129 (2016).
18. Gagnon, J. A. *et al.* Efficient Mutagenesis by Cas9 Protein-Mediated Oligonucleotide Insertion and Large-Scale Assessment of Single-Guide RNAs. *PLoS One* **9,** e98186 (2014).
19. Hwang, W. Y. *et al.* Heritable and Precise Zebrafish Genome Editing Using a CRISPR-Cas System. *PLoS One* **8,** e68708 (2013).
20. Bedell, V. M. *et al.* In vivo genome editing using a high-efficiency TALEN system. *Nature* **491,** 114–118 (2012).
21. Hoshijima, K., Jurynec, M. J. & Grunwald, D. J. Precise Editing of the Zebrafish Genome Made Simple and Efficient. *Dev. Cell* **36,** 654–667 (2016).

22. Bedell, V. M. & Ekker, S. C. Using Engineered Endonucleases to Create Knockout and Knockin Zebrafish Models. in *Chromosomal Mutagenesis: Second Edition* 291–305 (2015).

23. Vartak, S. V. & Raghavan, S. C. Inhibition of nonhomologous end joining to increase the specificity of CRISPR/Cas9 genome editing. *FEBS J.* **282,** 4289–4294 (2015).

24. Song, J. *et al.* RS-1 enhances CRISPR/Cas9- and TALEN-mediated knock-in efficiency. *Nat. Commun.* **7,** 10548 (2016).

25. Chu, V. T. *et al.* Increasing the efficiency of homology-directed repair for CRISPR-Cas9-induced precise gene editing in mammalian cells. *Nat. Biotechnol.* **33,** 543–548 (2015).

26. Maruyama, T. *et al.* Increasing the efficiency of precise genome editing with CRISPR-Cas9 by inhibition of nonhomologous end joining. *Nat. Biotechnol.* **33,** 538–542 (2015).

27. Cornet, C., Di Donato, V. & Terriente, J. Combining Zebrafish and CRISPR/Cas9: Toward a More Efficient Drug Discovery Pipeline. *Front. Pharmacol.* **9,** 1–11 (2018).

28. Vakulskas, C. A. *et al.* A high-fidelity Cas9 mutant delivered as a ribonucleoprotein complex enables efficient gene editing in human hematopoietic stem and progenitor cells. *Nat. Med.* **24,** 1216–1224 (2018).

29. Ota, S. *et al.* Efficient identification of TALEN-mediated genome modifications using heteroduplex mobility assays. *Genes to Cells* **18,** 450–458 (2013).

30. Miura, H., Quadros, R. M., Gurumurthy, C. B. & Ohtsuka, M. Easi-CRISPR for creating knock-in and conditional knockout mouse models using long ssDNA donors. *Nat. Protoc.* **13,** 195–215 (2018).

31. Richardson, C. D., Ray, G. J., DeWitt, M. A., Curie, G. L. & Corn, J. E. Enhancing homology-directed genome editing by catalytically active and inactive CRISPR-Cas9

using asymmetric donor DNA. *Nat. Biotechnol.* **34,** 339–344 (2016).

32. Symington, L. S. Mechanism and regulation of DNA end resection in eukaryotes. *Crit. Rev. Biochem. Mol. Biol.* **51,** 195–212 (2016).

33. Armstrong, G. A. B. *et al.* Homology Directed Knockin of Point Mutations in the Zebrafish tardbp and fus Genes in ALS Using the CRISPR/Cas9 System. *PLoS One* **11,** e0150188 (2016).

34. Bedell, V. M. & Ekker, S. C. Using Engineered Endonucleases to Create Knockout and Knockin Zebrafish Models. *Sci. Rep.* **6,** 291–305 (2015).

35. Inoue, D. & Wittbrodt, J. One for All—A Highly Efficient and Versatile Method for Fluorescent Immunostaining in Fish Embryos. *PLoS One* **6,** e19713 (2011).

36. Thisse, C. & Thisse, B. High-resolution in situ hybridization to whole-mount zebrafish embryos. *Nat. Protoc.* **3,** 59–69 (2008).

Chapter 4.

Conclusions

This dissertation is a method development study for a molecular analysis particularly for ChIP-seq of Sox proteins during the early embryonic development of zebrafish (*Danio rerio*). Mainly, methods related to CRISPR-mediated endogenous epitope tagging of *sox3* gene were developed. During the first part of this thesis project, a quite simple and relatively inexpensive approach was established to determine antibody affinity under IP conditions termed HiBiT-qIP. By using this method, one can predict in advance the performance of antibodies quantitatively under IP/ChIP assay conditions. The apparent affinities of interactions between five epitope tags, namely, FLAG, HA, PA, V5 and Ty1 and their cognate antibodies were determined. For some epitope tags, several antibody clones were examined for their affinity depending on the commercial availability. Most of the tag/antibody combinations showed high affinity. The use of epitope tags in multimeric form such as dimeric or trimeric form revealed a copy-number dependent increase in the apparent affinity and that improved the IP recovery significantly.

During the second part of this thesis project, an efficient epitope tag knock-in using CRISPR/Cas9 and long ssDNA donor in zebrafish was achieved and thereby, transgenic zebrafish lines were established with epitope (FLAGx3, PAx3) tagged Sox3 proteins. A significantly high efficiency of precise and heritable integration of the composite of epitope tags (~200 bp in length) to the *sox3* locus was achieved. In this work, we took advantage of the latest developments and/or technical advancements in the CRISPR/Cas9 tool kit. First, the use of protein-based Cas9 instead of Cas9 mRNA can facilitate early integration events. Engineered *S. pyogenes* Cas9 protein that has high fidelity and reduced off-target effects was used. Second, the use of synthetic crRNA and tracrRNA (two separate strands to form a complete guide RNA) may increase the cleavage efficiency. Here, chemically-modified and

length optimized variants of the native guide RNAs (Alt-R™ CRISPR crRNAs and tracrRNA, Integrated DNA Technologies, Coralville, IA, USA) were used. These improved CRISPR components must have contributed considerably to the highly efficient knock-in reported in this study. As a rule of thumb, achieving high rates of initial double-strand breaks through careful selection of highly efficient crRNAs is a prerequisite to increase the precise integration rates. Moreover, most importantly, the use of lssDNA and its effective structure were found to be a critical factor of the overall knock-in efficiency. As conclusive remarks, a shorter 3' homology arm preferably nearly 50-nts, with a longer 5' homology arm would be a better design for the lssDNA donor at least for an insert of about 200-nts.

As the future directions, the established transgenic fish can be further crossed to obtain stable genetically modified fish and can be used to perform molecular analyses such as ChIP-seq and/or functional proteomics analyses using LC-MS/MS (Liquid chromatography with mass spectrometry) by fully taking advantage of the knocked-in tags.

List of works

Publications

- **Ranawakage D.C.**, Takada T., Kamachi Y. HiBiT-qIP, HiBiT-based quantitative immunoprecipitation, facilitates the determination of antibody affinity under immunoprecipitation conditions, *Scientific Reports*, 9 (1): 6895 (2019)

- **Ranawakage D.C.** *et al.*, Efficient epitope tag knock-in using CRISPR-Cas9 and long ssDNA donor in Zebrafish, (in preparation)

Presentations

- **Ranawakage D.C.**, Takada T., Kamachi Y. HiBiT-qIP: HiBiT-based quantitative immunoprecipitation; an assay to help improve the success of IP/ChIP, British Society for Cell Biology and Developmental Biology spring meeting (2019), University of Warwick, UK (Poster)

- **Ranawakage D.C.**, Takada T., Kamachi Y. HiBiT-qIP, HiBiT-based quantitative immunoprecipitation, facilitates the determination of antibody affinity under immunoprecipitation conditions, Cell Biology and Developmental Biology joint meeting (2018), Tokyo, Japan (Poster)

- **Ranawakage D.C.**, Kamachi Y. Optimization of epitope tagging for Chromatin Immunoprecipitation, International Symposium on Frontier Technology (ISFT) (2017), Kochi, Japan (Oral, Conference proceeding full paper)

Acknowledgement

I am very much indebted to a host of amazing people who had helped me in various ways towards bringing out this dissertation, to all of whom I owe a deep sense of gratitude and appreciation. First and foremost, I wish to express my most profound gratitude to my supervisor, Professor Yusuke Kamachi, for giving me the opportunity to carry on this research in his lab, and for believing in myself and guiding me towards the end of this journey. I would like to express my sincere gratitude for my co-supervisors; Professor Takeshi Ohama and Associate Professor Sakae Horisawa for their advices and warm encouragement through-out these years. Also, I gratefully acknowledge Professor Seiji Tanaka and Professor Ishii Kojiro for their insightful comments and suggestions. Moreover, I would like to be thankful to the current and former members of the "*Kamachi lab* " for helping me in many ways.

It would have been difficult or impossible for me to witness this feeling of achievement unless my MSc supervisor, Professor Taku Okazaki gave me the opportunity to join his lab. Therefore, my sincere gratitude for "*Okazaki lab*" for opening the door of my career as a researcher. Moreover, I would like to be thankful to Professor Shinichiro Sakikawa for his enormous support and advices throughout my stay in Japan, and for showing me the life outside of the lab. Also, to the members of the international relations center for their immense support. Further, I gratefully acknowledge Special Scholarship Program (SSP) of Kochi University of Technology and Rotary Yoneyama memorial foundation scholarship for the financial support during this study. Also, I owe my deepest gratitude to my beloved parents, parents-in-law and my siblings for being always there for me and to have faith in me. And finally, last but by no means least, many thanks to my loving husband–my best friend and my mentor: Dr. E.K.C. Pradeep for always being by my side and for all of the sacrifices that you have made on my behalf.